

## **Design and Pilot of A Polygraph Field Validation Study**

**Thomas E. Dohm, Ph.D., and William G. Iacono, Ph.D.  
Personnel Decisions Research Intitutes, Inc.**

**July 1993**

**Department of Defense Polygraph Institute  
Fort McClellan, Alabama 36205-5114  
Telephone: 205-848-3803  
FAX: 205-848-5332**

**19960326 000**

**DTIC QUALITY INSPECTED 1**

**DISTRIBUTION STATEMENT A**

**Approved for public release;  
Distribution Unlimited**

*Enclosure b*

# REPORT DOCUMENTATION PAGE

*Form Approved  
OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)			2. REPORT DATE		3. REPORT TYPE AND DATES COVERED	
			July 1993		Final Report (From Sep 88 - Jul 93)	
4. TITLE AND SUBTITLE			5. FUNDING NUMBERS			
Design and Pilot of a Polygraph Field Validation Study			DoDPI88-P-0004			
6. AUTHOR(S)						
Thomas E. Dohm and William Iacono						
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)			8. PERFORMING ORGANIZATION REPORT NUMBER			
Personnel Decisions Research Institutes, Inc. 43 Main Street SE, Suite 405 Minneapolis, MN 55414			DABT02-88-C-0088			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER			
Department of Defense Polygraph Institute Building 3165 Fort McClellan, AL 36205-5114			DoDPI88-P-0004 DoDPI93-R-0006			
11. SUPPLEMENTARY NOTES						
12a. DISTRIBUTION/AVAILABILITY STATEMENT				12b. DISTRIBUTION CODE		
Public release, distribution unlimited						
13. ABSTRACT (Maximum 200 words)  This study was conducted to determine whether panel of individuals could accurately determine ground truth by reviewing case material contained in the files of individuals who underwent confession-verified polygraph tests. Panels were presented with all case material collected prior to the conduct of the polygraph exam. To determine how well panels could decide ground truth, the reliability with which panelists arrived at decisions of suspect-guilt and innocence was determined, and the validity of their decision was evaluated by tests. The study was based on selected specific-incident criminal investigative files requested from the military. Files were sanitized (removing all identifying information) with case material placed in chronological order and abstracted. Twenty-seven confession-verified multiple suspect criminal investigative files were then selected and reviewed by three six-member panels consisting of lay persons, police investigators, and attorneys. The results indicated that panelists, regardless of their professional backgrounds, could not decide which individuals committed crimes with high reliability. The decisions of the panelists were not accurate: their decisions did not match the ground truth criterion. These results indicated that it is unlikely that panels can be used to establish ground truth with any degree of confidence. Field studies of polygraph validity must use some other procedure to establish ground truth.						
14. SUBJECT TERMS				15. NUMBER OF PAGES 218		
panel study, validity, ground truth, confession-verified polygraph tests, reliability, field polygraph testing				16. PRICE CODE		
17. SECURITY CLASSIFICATION OF REPORT		18. SECURITY CLASSIFICATION OF THIS PAGE		19. SECURITY CLASSIFICATION OF ABSTRACT		20. LIMITATION OF ABSTRACT
Unclassified		Unclassified		Unclassified		

Report No. DoDPI93-R-0006

Design and Pilot of a Polygraph  
Field Validation Study

Thomas E. Dohm, Ph.D., and William G. Iacono, Ph.D.

July 1993

Department of Defense Polygraph Institute  
Fort McClellan, AL 36205

### Director's Foreword

Historically and ever increasingly, there is debate about the value of research results obtained in the laboratory setting and the generalization of those results to field application. When performing studies related to psychophysiological detection of deception (PDD), the academic, scientific and PDD practitioner debate focuses on the problem of determining "ground truth."

Published literature suggests expert panels should be able to accurately determine ground truth by reviewing case materials in the files of individuals who underwent confession-verified polygraph tests. The panel decisions would be based on investigative information absent the results of any PDD examination.

During this study, selected confession verified, multiple suspect, specific-incident criminal investigative files were reviewed by three six-member panels of lay persons, police investigators, and attorneys to determine how well they could decide ground truth, the reliability of their decisions of suspect guilt and innocence, and the validity of their decisions based on ground truth established by confessions after PDD examinations. The results reflect that panelists could not reliably or accurately determine which individuals committed the crimes in question. These results indicate that it is unlikely that panels can be used to establish ground truth with any degree of confidence and further, field studies of PDD validity must use some other procedure to establish ground truth.



Michael H. Capps  
Director

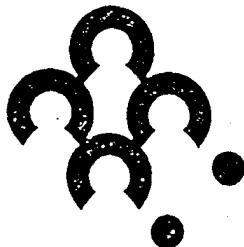
**DESIGN AND PILOT OF A  
POLYGRAPH FIELD VALIDATION STUDY**

Thomas E. Dohm and William G. Iacono

with literature review by

Cheryl Paullin and Kenneth T. Bruskiewicz

**July 1993  
Final Report  
Institute Report No. 227**



©Personnel Decisions Research Institutes, Inc.  
43 Main Street SE, Suite 405  
Minneapolis, MN 55414  
(612) 331-3680

## Foreword

A literature review focusing on methodological issues relevant to designing a field validation study and establishing measures of ground truth was completed. Since the proposed field validation study is to be conducted across all four military branches (i.e., Air Force, Army, Marines, and Navy), special attention was given to provide detailed information regarding current military procedure, polygraph methods employed within the military, and polygraph data storage and retrieval techniques.

Additional analyses were conducted separately using each military branch's computer archival data. Analysis of variance and chi-square tests were performed ascertain relationships between confession results/polygraph confirmation status with a series of demographic variables.

The major part of this report concerns a study that was conducted to determine whether panels of individuals could accurately determine grand truth by reviewing case material contained in the files of individuals who underwent confession-verified polygraph tests. Panels were presented with all case material collected prior to the conduct of the polygraph exam. To determine how well panels could decide ground truth, the reliability with which panelists arrived at decisions of suspect guilt and innocence was determined, and the validity of their decision was evaluated by comparing it to ground truth established from the confessions that followed failed polygraph tests. The study was based on selected specific-incident criminal investigative files requested from the military. Files were "sanitized" (removing all identifying information) with case material placed in chronological order and abstracted. Twenty-seven confession-verified multiple suspect criminal investigative files were then selected and reviewed by three six-member panels consisting of lay persons, police investigators, and attorneys. The results indicated that panelists, regardless of their professional backgrounds, could not decide which individuals committed crimes with high reliability. Furthermore, the decisions of the panelists were not accurate: Their decisions did not match the ground truth criterion.

These results indicate that it is unlikely that panels can be used to establish ground truth with any degree of confidence. Field studies of polygraph validity must therefore use some other procedure to establish ground truth. Recommendations regarding how this might be accomplished are presented in the second part of this report.

This research was sponsored by the Department of Defense Polygraph Institute (DoDPI) project DoDPI88-P-0004, under a contract administered by the Directorate of Contracting, United States Department of the Army (contract DABT02-88-C-0088). The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government. Appendices in another volume.

## Table of Contents

<b>CONCEPTUAL OVERVIEW .....</b>	1
<b>LITERATURE REVIEW .....</b>	5
Notable Figures in the History of the Polygraph .....	5
Polygraph Methodology .....	7
Types of Polygraph Procedures .....	8
Relevant/Irrelevant (R/I) Technique .....	8
Guilty Knowledge Test .....	8
Control Question Technique .....	9
Operational Definitions .....	9
Past Attempts at Establishing the Validity of the Polygraph .....	12
Laboratory vs. Field Designs .....	12
Past Field Validation Research .....	13
Methodological Concerns .....	14
Establishing Ground Truth .....	14
Individual Differences in Detection of Deception .....	23
Amount of Information Available to Examiner .....	23
Use of Countermeasures .....	24
Base Rate Information .....	25
Good Validation Study Procedures .....	25
<b>GENERAL PROCEDURES FOR THE USE OF THE POLYGRAPH IN MILITARY SETTINGS .....</b>	27
Some Differences Between Military and Civilian Settings .....	27
General Use of the Polygraph in the Military .....	27
Polygraph Procedures: Specific Comparisons Between Branches.....	29
Applicability .....	29
Uses .....	29
Authorization (for Approval) .....	30

## Table of Contents (continued)

Authorization (Forms and Criteria) .....	30
Administration Procedure .....	31
Computer File Maintenance of Polygraph Data .....	34
Analyses of Military Computer Files .....	37
Frequency and Cross-Tabulation of Crime Category by Year Military Branch .....	38
Polygraph Results of Military Branch.....	38
Selection of Case Files .....	46
Creating a Taxonomic Structure of Criminal Investigative Cases .....	46
Initial Selection of Case Files to Review .....	46
Final Selection of Case Files to Review.....	54
Rules for "Sanitizing" Files .....	56
Procedure for Abstracting Polygraph Files .....	59
Procedure .....	59
Case Handling Procedures .....	64
Case Log .....	66
Missing Case Files .....	66
Data Analyses - Archival Data.....	72
Air Force .....	74
Army.....	77
Marines .....	84
Navy .....	88
<b>PANEL STUDY .....</b>	<b>98</b>
Description of Cases Used .....	101
Selection of Cases for Panels .....	102
Description Rating Forms.....	105

## Table of Contents (continued)

Description of Cases Used.....	109
Data Collection.....	112
Panel Study: Statistical Analyses and Results.....	112
Interrater Reliability of Panel Ratings.....	115
Reliability of Specific Cases .....	130
Global Rating Changes.....	132
Leniency of Ratings.....	140
Relationship Between Specific Questions to Global Ratings .....	144
Accuracy of Outcome.....	144
What is the Overall Accuracy of the Ratings .....	146
What is the Overall Accuracy of the Ratings for Each Panel .....	148
What is the Overall Accuracy of the Ratings for Each Suspect.....	148
What is the Overall Accuracy of the Ratings for Each Suspect for Each Panel.....	157
What is the Overall Accuracy of the Ratings for Each Panel Member.....	157
What is the Overall Accuracy of the Specific Ratings .....	159
Comparison of Polygraph Results .....	164
EXECUTIVE SUMMARY .....	169
Data Collection Procedure.....	174
Review of Cases by Panel .....	175
Results .....	176
Accuracy of Outcome.....	177
Summary of Panel Study Results .....	178
Did the Type of Panel Make a Difference? .....	179

## Table of Contents (continued)

What is the Optimal Size for a Panel?.....	180
Are Confessions Suitable for Establishing Ground Truth.....	182
Will Better Selection or Training of Panel Members Produce "Better" Results? .....	182
Recommendations for Future Research .....	182
Use of Military Records .....	183
Access to Data Files .....	183
Standardization of Military Polygraph Procedure .....	184
Standardization of Computer Databases .....	186
Future Field Validity Studies.....	188
Redo the Panel Study.....	188
New Designs for Field Validation of the Polygraph .....	193
Design 1.....	194
Design 2.....	196
References .....	199

## Appendices

A: Air Force Polygraph Procedures .....	A-1
B: Army Polygraph Procedures .....	B-1
C: Navy and Marine Polygraph Procedures .....	C-1
D: Investigative Categories Across Military Branches .....	D-1
E: Air Force Polygraph Data Cross-Tabulation OF Criminal Code by Year .....	E-1
F: Army Polygraph Data Cross-Tabulation of Criminal Code by Year .....	F-1
G: Marine Polygraph Data Cross-Tabulation of Criminal Code by Year .....	G-1
H: Navy Polygraph Data Cross-Tabulation of Criminal Code by Year.....	H-1
I: Polygraph Results by Confirmation Status: Air Force (1986-1989) .....	I-1
J: Polygraph Results by Confirmation Status: Army (1983 - 1989) .....	J-1

K: Polygraph Results by Confirmation Status: Marines (1987-1989) .....	K-1
L: Polygraph Results by Confirmation Status: Navy (1980-1987).....	L-1
M: Case Log .....	M-1
N: Data Analyses of Air Force Archival Data.....	N-1

## Appendices (continued)

O: Data Analyses of Army Archival Data.....	O-1
P: Data Analyses of Marine Archival Data .....	P-1
Q: Data Analyses of Navy Archival Data .....	Q-1
R: Panel Study: Demographic Questionnaires .....	R-1
S: Case-Specific Rating Forms Used During Panel Review .....	S-1
T: Listing of Specific Items which Significantly Correlate with Global Ratings for Each Case Across All Panel Member.....	T-1
U: Accuracy of Outcome Data .....	U-1
V: Accuracy of Outcome: Specific Ratings for Each Question for Each Panel Ground Truth Hit Rates for Specific Questions .....	V-1
W: Hit Rates: Polygraph Results by Ratings for Specific Questions.....	W-1

## List of Tables

1. Selected Polygraph Validation Studies with Various Methodological Consideration .....	16-21
2. Information Available on the Computer Polygraph Files by Branch.....	35-36
3. Frequency of Crime by Criminal Category (Air Force Data: 1987-1988).....	39
4. Frequency of Crime by Criminal Category .....	40-41
5. Frequency of Crime by Criminal Category (Marine Data: 1987-1989).....	42
6. Frequency of Crime by Criminal Category (Navy Data: 1980-1987).....	43-45
7a. Polygraph Result by Confirmation Status Across All Branches - 1983.....	47
7b. Polygraph Result by Confirmation Status Across All Branches - 1984.....	48
7c. Polygraph Result by Confirmation Status Across All Branches - 1985.....	49
7d. Polygraph Result by Confirmation Status Across All Branches - 1986.....	50
7e. Polygraph Result by Confirmation Status Across All Branches - 1987.....	51
7f. Polygraph Result by Confirmation Status Across All Branches - 1988.....	52

## List of Tables (continued)

7g. Polygraph Result by Confirmation Status Across All Branches - 1989 .....	53
8. Air Force Data: Crosstabulation of Relevant Variables with Polygraph Results.....	75
9. Air Force Data: Crosstabulation of Relevant Variables with Confession.....	78
10. Army Data: Crosstabulation of Relevant Variables with Polygraph Results.....	79
11. Army Data: Crosstabulation of Relevant Variables with Confession.....	81
12. Army Data: One-way Analysis of Variance: Length of Polygraph Exam (in Minutes) by Relevant Variables.....	83
13. Marine Data: Crosstabulation of Relevant Variables with Polygraph Results .....	85
14. Marine Data: Crosstabulation of Relevant Variables with Confession Results .....	87
15. Navy Data: Crosstabulation of Relevant Variables with Polygraph Results .....	89
16. Navy Data: Crosstabulation of Relevant Variables with Confession.....	91
17. Navy Data: One-way Analysis of Variance: Length of Polygraph Exam by (in Minutes) by Relevant Variables.....	93
18. Navy Data: One-way Analysis of Variance: Length of Post-Test Exam by Relevant Variables.....	95
19. Navy Data: One-way Analysis of Variance: Length of Chart Time by Relevant Variables.....	97
20. Panel Composition: Gender, Age, Race, Education, and Experience.....	103
21. Panel Case Descriptions: Number of Pages, Suspects, Crimes, and Ratings .....	110-111
22. Average Composite Reliability of Specific Questions.....	118
23. Number of Reliability Estimates by Panel Size .....	119
24. Overall Reliabilities - Corrected for Case All Combinations of 6 at a Time .....	120
25. Overall Reliabilities - Corrected for Case All Combinations of 5 at a Time .....	121
26. Overall Reliabilities - Corrected for Case All Combinations of 4 at a Time .....	122
27. Overall Reliabilities - Corrected for Case All Combinations of 3 at a Time .....	123
28. Overall Reliabilities - Corrected for Case All Combinations of 2 at a Time .....	124

## List of Tables (continue)

29. Average Composite Reliabilities of Global Ratings by Case (All Combinations) of 6 at a Time.....	131
30. Mean Global Ratings by Suspect/Victim .....	133-139
31. Initial Versus Final Global Ratings (Lay Panel) .....	141
32. Initial Versus Final Global Ratings (Police Panel) .....	142
33. Initial Versus Final Global Ratings (Lawyer Panel) .....	143
34. Accuracy of Outcome: Global Ratings of Panel with Independent Judges .....	147
35. Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim of Each Case .....	149-156
36. Accuracy of Ratings for Each Panel Member .....	158
37. Accuracy of Outcome: Global and Specific Ratings for Combined Sample .....	160-163
38. Hit Rates: Global Ratings of Panel with Polygraph Results by Panel Type .....	165
39. Hit Rates: Global Ratings with Polygraph Results by Panel Member and by Case Number .....	167
40. Hit Rates: Polygraph Results with Independent Judges Estimates of Ground Truth.....	168
41. Power Analysis of Global 2 Ratings .....	169
42. Overall Findings .....	181

## Figures

1. Contingency Table for Reporting Polygraph Results.....	11
2. Abstract Summary Form .....	67-68
3a. Quality of Confession Rating Scale .....	71
3b. Final "Quality of Confession Rating Scale" .....	73
4a. Envelope 1 - Global rating 1; Case Number 1 .....	106
4b. Envelope 1 - Initial rating; Case Number 1 .....	107

## **Figures (continued)**

5. Case Review Process .....	113-114
6. Average Composite Reliability of Initial and Final Ratings by Size of Layperson Panel .....	126
7. Average Composite Reliability of Initial and Final Ratings by Size of Police Panel .....	127
8. Average Composite Reliability of Initial and Final Ratings by Size of Lawyer Panel .....	128
9. Average Composite Reliability of Initial and Final Ratings by Size of Combined Panel.....	129

### **Acknowledgements**

This technical report was prepared for the Polygraph Institute under Contract No. DABT02-88-C-0088.

A project of this scope requires the cooperation and assistance from many organizations and individuals. With the full understanding that the requirements of our project placed additional demands on the criminal investigative and polygraph units of each military branch's workload, we are very appreciative of the high level of cooperation given by each of these groups in helping us carry out our research goals. We are also very thankful of the Polygraph Institute's assistance in providing the additional human resources necessary to complete the process of sanitizing the files.

In particular, we would like to thank the following individuals who were instrumental in coordinating our data collection requests and providing us with the necessary background materials:

For the Air Force, LTC Joseph Christ, Gale Ahern, Cecil Fry, and James Morrison;

For the Army, Robert A. Brisentine, Wilbur Hardy, Wolfgang Vinsky, and Gary Light;

For the Marines: Major Lance Mueller and Michelle S. Schilling; and

For the Navy: David J. Karpovich and Denise Stakes-Bugsey.

Finally, we would like to thank Dr. Gordon Barland who facilitated our requests for additional personnel support to complete the major undertaking of sanitizing the criminal investigative files. Dr. Barland's valuable guidance and efficient handling of numerous requests for information served us well throughout the project.

## CONCEPTUAL OVERVIEW

Although there have been a number of studies that have endeavored to assess the validity of field polygraph testing, no consensus has yet been reached on the accuracy of polygraph testing. Recent reviews (Iacono, 1989; Iacono & Patrick, 1987, 1988), concluded that this lack of consensus does not simply reflect the prejudices of those reviewing these studies, but rather, that it stems from the fact that these studies are too fraught with methodological and reporting deficiencies to provide convincing answers to the accuracy question.

Field validity studies conducted to date have used as the criterion of ground truth either complete or partial panel consensus or confessions from the accused. Both types of studies have serious inherent limitations.

Confession studies have been criticized because confessors are a select group. While this is true, a more serious problem with confession studies is that the polygraph charts associated with confessions represent a select group (Iacono, 1989, Iacono & Patrick, 1987, 1988). In the usual case, confessions are obtained when the examiner attempts to elicit them from a subject whose charts are indicative of deception. Thus, confessions will be obtained only from those who produced deceptive charts. The charts associated with these "false negative" errors will never become part of a validity study when confessions are used to establish ground truth. Likewise, an unknown number of innocent people will erroneously produce deceptive charts. However, these individuals will generally not confess to an act they did not commit, and these "false positive" errors will never become part of the validity study. Note that if some innocent people do confess following a failed polygraph test, this type of error is incorrectly counted as a hit in validity studies using confessions.

To summarize, the major problem with confession studies is that almost all of the cases wherein the original examiner made an error are systematically excluded from the data sample. The only

cases included are those where both the charts indicate deception and the examinee confesses. Since numerical scoring is highly reliable, giving these charts of confession verified guilty and innocent subjects to other examiners to rescore blindly will of course confirm the original examiner's assessment. As a consequence, in confession studies, because cases where potential errors could have been made are eliminated, inaccurate estimates of polygraph validity will be obtained. Accuracy rates for guilty individuals will be especially overly optimistic and should be close to 100% when numerical scoring is used even if the polygraph test had zero validity. Interpreting the data from innocent subjects will also be misleading for these same reasons (see Iacono, 1989, for a full discussion of the consequences of using confessions as the ground truth criterion).

It is of special significance to note that the arguments developed above also apply to panel studies in which the panels have access to confessions elicited following failed polygraph tests. A confession is likely to be persuasive evidence to a panel. However, because confessions will be associated with an unrepresentative set of charts, confessions elicited following failed examinations should not be presented to the panel if an accurate estimate of polygraph validity is desired.

In the Honts, Raskin, Kircher, and Horowitz (1988) confession study, confessions were corroborated by additional evidence indicating that the confessor was in fact guilty. Hence, after confession to a crime following a failed polygraph test, the accused might also turn over stolen items, etc. Such corroborating information, because it is also associated with an unrepresentative set of charts, must also be denied the panel if accurate validity estimates are the goal. Confirmatory data that is uncovered independent of the confession can and should be presented to the panel.

A limitation to both is that the ground truth criterion can be obtained in only a small number of cases. In most instances, polygraph tests are given when there is insufficient evidence to prosecute or otherwise resolve a case. Under these circumstances, the information that can be pre-

sented to a panel will also be ambiguous, making it likely that the panel will be unable to reach conclusions with confidence. By the same token, individuals who confess to crimes also represent a select group of the entire population of individuals who undergo polygraph tests. Hence, both panel and confession studies can only resolve ground truth for a fraction of those taking polygraph tests. The extent to which these fractions are representative of all polygraph cases is an open question.

Panel studies conducted to date are also problematic in that it is unclear how much and what types and quality of data the panels had to work with. It is imperative that the data be as complete and accurate as possible. In this regard, it is important that the material presented to the panel, if it is abstracted, be gathered with demonstrated reliability so that pertinent facts are not "filtered" by some person's idiosyncratic views of what is and what is not important. In addition, of special concern is whether or not panels have access to confessions elicited following a failed polygraph test. To the extent that they do have these confessions, the panel will most likely affirm the confession, and the distinctness of these two methods for establishing ground truth becomes blurred.

The present research project provides a unique opportunity to more systematically examine the potential of panel studies. Panel studies that have been conducted to date have the methodological limitations described above, so their potential as a means for establishing ground truth has not been fully explored. The research requirements calls for a pilot study in which several issues will be assessed. The first deals with the possibility that the composition of the criterion panel may influence the accuracy of the panel. Examinations of this issue will also make it possible to determine how best to constitute criterion panels in future research. Another aspect is the level of information provided to the panel. Can a panel be used at on earlier stages of an investigation (i.e., prior to the administration of the polygraph) and still have adequate accuracy? A final aspect of the pilot study concerns a) how accurate the panel decisions are vis a vis a confession as the criterion of ground truth, and b) how believable confessions are to a panel that considers

them carefully along with other evidence.

It should be apparent from the above that a field validity study cannot accurately assess polygraph test validity if the only ground truth criterion is post-polygraph test confession. The results of the panel study discussed above will determine the adequacy of this approach for resolving ground truth.

To further clarify the issues involved in conducting a field validation study in a military setting, the remainder of this report will present:

- 1) a brief history of the polygraph, polygraph techniques, and operational definitions
- 2) a review of the literature focussing on methodological issues
- 3) a discussion of the use of the polygraph in military settings with specific comparisons between branches
- 4) a discussion of the computer database management (by the various military branches) of polygraph case information
- 5) procedures and methods for selecting, organizing, and indexing military criminal case files
- 6) analyses of archival polygraph data contained in the computer files of the various branches. Special emphasis will be made to analyze the relationship between the results of the polygraph exam and its confirmation status with a series of demographic variables
- 7) the results of the panel study. Independent variables were size of panel, (two to six members), type of panel (lay, police, and lawyer), and the level of information provided to the panel. The reliability of the panel ratings and the accuracy of these ratings are fully discussed relative to design issues
- 8) a series of recommendations for conducting polygraph validation research in a military setting.

## LITERATURE REVIEW

(Cheryl Paullin, Kenneth T. Bruskiewicz, and Thomas E. Dohm)

### Notable Figures in the History of the Polygraph

William Marston was one of the most successful of the early polygraphers, as well as one of the most vocal. Although he was not the first to apply experimental principles to the detection of deception, Marston developed a procedure in which he used a sphygmomanometer to measure a suspect's blood pressure as the suspect answered questions pertaining to some alleged crime (Abrams, 1977). Marston was able to discern whether or not the suspect was lying by assessing the suspect's blood pressure. Between 1913 and 1923, Marston reported that he made accurate judgments in over 95% of his polygraph administrations (Abrams, 1977).

John Larson further developed the polygraph procedure by combining the measurement of blood pressure with measures of respiratory changes during questioning (Lykken, 1984). Larson is credited with assembling the first continuous-recording polygraph and, like Marston, also claimed great success in the application of the instrument. Larson required his subjects to respond "yes" or "no" to direct questions (Abrams, 1977). Some of the questions were related to the crime under investigation and some were not. This technique formed the foundation for what later became known as the Relevant/Irrelevant technique. In one incident Larson used his technique to investigate a case of shoplifting from a local store. A store clerk knew only that the shoplifter was one of the thirty-eight women living in a nearby dormitory. Larson examined all thirty-eight of the suspects and judged thirty-seven to be innocent and one deceptive. The one suspect deemed deceptive subsequently confessed to the crime and the case was solved.

Another notable figure in the development of the polygraph was Leonarde Keeler. Keeler began his career as a high school assistant to Larson (Lykken, 1984). Keeler developed one of the first

portable field polygraphs and later formed the first polygraph training school in the United States. He expanded polygraph technology by introducing the Peak-of-Tension Test (POTT) and the Stimulation Test. The POTT is used when the polygrapher has information about a crime that is possessed by only the guilty party and the investigators. The suspect is asked a series of questions, with each series concerning a critical detail related to the crime. In each series the non-critical questions are the same, but the critical item is different (one critical item and six or seven non-critical distractors). Peak-of-Tension refers to the suspect's physiological arousal which occurs in anticipation of lying to the critical item (Iacono & Patrick, 1988). One advantage of the POTT is that it is not likely to result in a large number of false positives (i.e., innocent persons judged deceptive) because innocent persons are not likely to have knowledge of the crime.

The Stimulation Test is seen by many as critical to the polygraph procedure (Elaad & Kleiner, 1986). In the Stimulation Test, a situation is set up in which the suspect is forced to lie. This allows the examiner to demonstrate how easily a lie can be detected. The Stimulation Test serves two purposes: it is intended to reduce the fears of innocent suspects while causing guilty suspects to be more reactive to the relevant questions.

Keeler was also one of the first to recognize that the term "lie detector" was a misnomer (Abrams, 1977). The polygraph does not detect lying; it simply records various physiological measures of the suspect which must be interpreted by the examiner. The examiner uses polygraph charts and other subjective information to make a judgment of deception indicated (DI) or no deception not indicated (NDI).

John Reid played an even larger role in the development of the polygraph than did Keeler (Lykken, 1984). He founded the Reid College of Detection of Deception, the first accredited school of polygraph permitted to award Master's degrees. Reid also developed the Control Question Technique (CQT). Briefly, the CQT contains three types of questions. Irrelevant ques-

tions are designed to help the subject relax and are not used in making a judgment of DI or NDI. Relevant and control questions are used to assess the subject's truthfulness and are critical for the polygraph procedure. Relevant questions are directly related to the crime under investigation and are intended to elicit stronger physiological arousal in guilty subjects than in innocent subjects. Control questions are designed to tap information about past incidents that are similar to the matter at hand. It is expected that innocent subjects will have stronger physiological reactions to the control questions than guilty subjects because the questions are constructed such that the content concerns the innocent subject directly (Iacono & Patrick, 1988).

C. Backster, a student of Reid's, introduced the "zone of comparison" format, which includes the comparison of responses to adjacent relevant and control questions (Iacono & Patrick, 1988). This allows for a semi-objective scoring of the polygraph charts. Although this format allows for more objectivity, it has not been universally accepted, and some polygraph examiners still make judgments using a global scoring technique.

### Polygraph Methodology

The polygraph technique refers to a relatively complex set of procedures for asking questions and measuring physiological responses in order to detect deception and/or establish truth (Office of Technology Assessment, 1983). Typically, three sets of physiological measures are taken during the polygraph examination. Galvanic Skin Response (GSR) is measured with electrodes attached to the fingertips. Cardiovascular measures are recorded using a standard blood pressure cuff attached to the upper arm. Respiration is measured via two pneumatic belts that are positioned around the upper thorax and abdomen. Other measures have also been used in conjunction with the polygraph (e.g., Janisse & Bradley, 1980), but skin conductance, respiratory, and cardiovascular measures are by far the most common. Some polygraph measures can be taken either electronically or mechanically, although it does not seem to matter which type of instrument is used (Iacono & Patrick, 1987). There is no characteristic "lie response". Detection of

deception judgments are made by the polygrapher who interprets the subject's physiological patterns of responding to the questions during the polygraph examination. The following paragraph briefly describes the most common techniques used by polygraphers in detecting deception.

### **Types of Polygraph Procedures**

#### **Relevant/Irrelevant (R/I) Technique**

Keeler's R/I technique is the predecessor of almost all current polygraph techniques (Abrams, 1989). Although the R/I technique was used frequently in criminal investigations prior to the development of the CQT, it is currently used primarily in pre-employment and employee screening procedures (Iacono & Patrick, 1988). The R/I technique consists of two types of questions. Relevant questions, pertain to the matter at hand and irrelevant questions serve as comparison measures. More recently, some polygraph examiners have utilized a Relevant/Relevant (R/R) technique in screening-type situations. Generally speaking, the R/R technique also consists of relevant questions and irrelevant questions. However, in this procedure the irrelevant questions are not used as comparators, they are simply used to establish a baseline and to help the subject to relax and feel more at ease. A subject's truthfulness is determined by comparing his or her physiological responses to the various relevant questions, over a number of different issues. If the subject reacts more strongly to certain relevant issues, as compared to other relevant issues, he or she is likely to be judged deceptive on the issue for which he or she showed the greatest arousal.

#### **Guilty Knowledge Test**

The guilty knowledge test (GKT), like the POTT, does not attempt to detect deception, rather, it attempts to detect whether the subject possesses knowledge of the crime that only the guilty party could process (Lykken, 1990). Use of the GKT depends upon the investigator gathering information about the case that is likely to be known only to the guilty party and criminal investigators. This information is presented to the subject in a multiple choice format and the polygraph is used to determine whether the subject reacts more strongly to items related to the crime

than to items not related to the crime. A drawback to this technique is that it may be quite difficult to maintain the confidentiality of the details needed for the GKT, although Lykken (1988) argues that a careful investigator should be able to gather the necessary information. The GKT has been used primarily in laboratory studies. Elaad (1990) was the first researcher to attempt to validate the GKT in a real-life criminal investigation. He was able to correctly classify 94% of the innocent subjects and 65% of the guilty subjects excluding inconclusives (i.e., those subjects the polygraph examiner was unable to classify as DI or NDI).

### **Control Question Technique (CQT)**

The CQT with the zone-of-comparison format is currently the most often used polygraph technique (Iacono & Patrick, 1987). The CQT test usually consists of approximately 10 questions. General questions are used to put the subject at ease and are not used in scoring the polygraph charts (e.g., "Is your name David?"). Relevant and control questions are used to evaluate the subject's truthfulness. Relevant questions are those that pertain directly to the crime under investigation (e.g., "Did you steal the night vision goggles?"). Control questions ask the subject about past behaviors that are similar to the main issue of the test (e.g., "Have you ever stolen something of value?"). According to CQT theory, the guilty party should be more physiologically aroused by the relevant questions than the control questions, because he/she was involved in the crime. The innocent party should be more physiologically aroused by the control questions than the relevant questions because he/she was not involved in the crime but has most likely committed the socially undesirable acts covered by the control questions. The CQT procedure is typically a very long and involved process that requires an extensive pretest interview with the subject in order to determine appropriate control questions, followed by the polygraph testing phase, and finally, if necessary a post-test interrogation. It is during the post-test interrogation that most polygraph-related confessions are obtained. Bersh (1969) was one of the first to conduct a field validation study of the polygraph using the CQT. He found accuracy rates, for example, of 80% for innocent subjects and 71% for guilty subjects when comparing the outcome of polygraph examinations to the majority judgment of a Judge Advocate General

(JAG) panel.

### **Operational Definitions**

One of the many problems in assessing the literature concerning the polygraph is the lack of a common terminology base to use to compare studies.

Below are definitions for some commonly used polygraph terminology.

**Deception Indicated (DI)**--refers to a polygraph test outcome in which the subject is judged to be deceptive (i.e., lying, or not forthright in responding to questions). This outcome is viewed as signaling subject guilt.

**No Deception Indicated (NDI)**--refers to a polygraph test outcome in which the subject is judged to be telling the truth (i.e., non-deceptive in responding to questions). This outcome is viewed as signaling subject lack of guilt (or innocence).

**False Positives (FPs)**--This is a situation in which an innocent subject is misjudged as DI. In statistical terms, this error is known as a Type I error.

**False Negatives (FNs)**--This is a situation in which a subject who is actually guilty is judged to be NDI, or not guilty. In statistical terms, this error is known as Type II error.

**Base Rate for Deception**--refers to the proportion of the population known to be lying to one or more of the relevant questions during the polygraph examination (Barland, 1988). Generally, this is expressed as a percentage.

Past validation studies (e.g., Bradley & Janisse, 1981) have reported overall accuracy rates for guilty and innocent subjects combined as indices of validity. Lykken (1977) argues that this is a misleading method of reporting accuracies. Specifically, accuracy rates (or hit rates) should be reported separately for both guilty and innocent subjects. Figure 1 shows possible combinations of criterion and test result. Innocent Hit Rate (IHR) is equal to  $D/(B + D)$ .

Figure 1

Contingency Table for Reporting Polygraph Results

		Criterion	
		Guilty	Not Guilty
Test Result	Guilty	A	B
	Not Guilty	C	D

$$IHR = D/(B+D)$$

$$GHR = A/(A+C)$$

$$CHR = (GHR+IHR)/2$$

Guilty Hit Rate (GHR) is equal to  $A/(A + C)$ . Corrected Hit Rate (CHR) is an estimate of what the hit rate would have been if there had been the same number of guilty subjects as innocent subjects and is equal to  $(GHR + IHR)/2$ . Lykken (1977) advocates reporting all three hit rates in any validation study of the polygraph. Barland (1988) also suggests reporting whether accuracy rates refer to original examiner decisions or examiner decisions based on a blind chart analysis. Most recent validation studies have followed these guidelines (e.g., Kircher & Raskin, 1988; Patrick & Iacono, 1991), although, as Patrick and Iacono note, it is difficult to use field studies to estimate polygraph accuracy for the guilty.

### **Past Attempts at Establishing the Validity of the Polygraph**

#### **Laboratory vs. Field Designs**

When discussing attempts to establish the validity of the polygraph, it is important to distinguish between field and laboratory validation study designs. Typically, a laboratory (or analog) study of the polygraph utilizes a mock-crime procedure in which subjects simulate committing a crime (e.g., Barland & Raskin, 1975, 1978; Bradley & Ainsworth, 1984; Podlesney & Raskin, 1977) and a polygraph test is conducted to see if the guilty subjects can be identified. A laboratory study has the advantage of establishing an unambiguous criterion of ground truth. Therefore, in laboratory studies it is possible to clearly establish a functional relationship between the polygraph test and detection of deception (i.e., internal validity). A serious problem with the use of laboratory studies is the issue of external validity, or the generalizability of the findings to the rest of the population of interest. Subjects participating in laboratory study may not be highly motivated to defeat the polygraph test. There is much less fear associated with being declared "deception indicated" (DI) in a laboratory situation than in a real-life criminal investigation. More recently, several studies (Ginton, Daie, Elaad, & Ben-Shakhar, 1982; Patrick & Iacono, 1989) have tried to overcome these problems by using procedures which attempt to simulate real-life crimes in a realistically threatening context. Ginton et al. (1982) outline the conditions necessary to validate the ability of the polygraph to detect deception: (a) the act must be authentic and freely undertaken, (b) it must be independently ascertainable which subjects actually

committed the act, (c) the subjects must believe that the interrogator does not know who committed the act, (d) the subjects must be genuinely concerned about the outcome of the polygraph test, (e) the polygrapher must have access only to the polygraph charts, and (f) the polygrapher must not know the proportion of guilty and innocent subjects in the sample. Ginton et al. (1982) were able to satisfy these conditions using a very unique experimental design (interested readers are referred to the article for details about the study). They accurately classified 68.8% of the guilty subjects and 36.5% of the innocent subjects (*including* inconclusives; 100% & 83%, respectively, excluding inconclusives) based on a field scoring technique conducted by eight polygraphers. Patrick and Iacono (1989) evaluated the accuracy of the CQT with psychopaths within a realistically threatening context and were able to correctly classify 87% of the guilty subjects and 56% of the innocent subjects (excluding inconclusives). These findings suggest that a situation in which polygraph subjects are genuinely concerned about the outcome of the polygraph test will most likely result in a high number of false positive errors. Despite these attempts to make the polygraph situation more realistic, laboratory studies still show limited generalizability. A field validation design is necessary to establish external validity.

### **Past Field Validation Research**

Iacono and Patrick (1987, 1988) reviewed all of the published field studies that attempted to assess the validity of the CQT in detecting deception. Among the deficiencies noted in their review are: (a) a failure to report the criteria used to establish ground truth (Hunter & Ash, 1973; Slowick & Buckley, 1975; Wicklander & Hunter, 1975; Yankee, Powell, & Newland, 1986); (b) providing insufficient detail pertaining to how polygraph records were sampled (Davidson, 1979; Horvath & Reid, 1971; Hunter & Ash, 1973; Wicklander & Hunter, 1975); (c) the exclusion of subjects from the sample using arbitrary criteria (Davidson, 1979; Horvath & Reid, 1971); and (d) failure to specify the procedure used to identify the initial sample of cases (Horvath & Reid, 1971; Hunter & Ash, 1973; Slowick & Buckley, 1975; Wicklander & Hunter, 1975). Iacono and Patrick (1988) also report that the hit rates for both innocent and guilty subjects are generally lower in studies conducted by researchers who are social scientists and

trained polygraphers, than in studies conducted by trained polygraphers who are not social scientists.

Generally speaking, the most methodologically sophisticated polygraph field validation studies have found high levels of accuracy for guilty subjects, but low levels of accuracy for innocent subjects (Barland & Raskin, 1976; Horvath, 1977; Patrick & Iacono, 1991), although, as Patrick and Iacono note, it is difficult to use field studies to estimate polygraph accuracy for the guilty.

Patrick and Iacono (1991) conducted the most recent, and perhaps the most methodologically sound, validation study of the CQT. They collected data for every polygraph test conducted by the Vancouver Polygraph Section of the Royal Canadian Mounted Police (RCMP) for its nine largest detachments between 1980 and 1984. For cases which could be verified with maximum certainty via information obtained *independently* of the polygraph ( $N=75$ ), an accuracy rate of 56% was attained for innocent subjects (excluding inconclusives). Patrick and Iacono (1991) also report that they were able to verify only 2% of the cases independently of the polygraph for the guilty subjects, contrasted with 65% of the cases for the innocent subjects. Hence, it was not possible to estimate the accuracy of the polygraph with guilty subjects. They suggest that the difference in accuracy rates in studies with confession-verified samples may be due to a dependency between the criterion and the polygraph charts. A judgment of DI in a polygraph examination provides reason for the examiner to pursue interrogation; if the suspect subsequently confesses, the outcome is confirmed. However, if a guilty subject is judged NDI, the investigation ends and the suspect's guilt is never discovered. Thus, the accuracy rate for the guilty subjects may be inflated because of this dependency. If this sampling bias were controlled for the accuracy rates for guilty subjects might be closer to that of the innocent subjects. These data underscore the need to verify, independently of the polygraph, as many cases as possible from any sample in a validation study.

Please refer to Table 1 for a summary review of select polygraph validation studies relative to several methodological considerations.

### **Methodological Concerns**

#### **Establishing Ground Truth**

Establishing ground truth is arguably the most problematic issue in assessing the validity of the polygraph. Briefly, establishing ground truth refers to the criterion used to establish a subject's guilt or innocence. In a polygraph validation study this criterion is compared to the outcome of the polygraph examination. The extent to which the outcome of the polygraph examination is in accordance with ground truth determines the validity of the polygraph in detecting deception. In past field validation studies of the polygraph, ground truth has been established via confessions (Davidson, 1979; Horvath, 1977; Horvath & Reid, 1971; Patrick & Iacono, 1991), panel judgments (Barland & Raskin, 1976; Bersh, 1969) and court outcomes (Barland & Raskin, 1976). Confession-verification is by far the most frequently utilized method of establishing ground truth. According to Iacono (1991) there are three serious concerns that arise when the criterion for ground truth is a confession. The first is that if the confession occurs prior to or during any part of the polygraph examination, it cannot be considered an independent measure of guilt. Another concern is that the person who confesses may not be guilty (i.e. an innocent person who realizes that they most likely will be convicted may confess in exchange for a lighter sentence). Finally, it is possible that those who confess may be a select sample of subjects, which may bias the outcome of the study.

The most serious threat to validity in polygraph studies in which ground truth is established with post-polygraph confessions is probably sampling bias (Patrick & Iacono, 1991). A judgment of Deception Indicated in a polygraph examination provides reason for the examiner to pursue interrogation. If the subject subsequently confesses, the polygraph outcome is considered verified. However, guilty subjects judged to be truthful are not interrogated and the polygraph outcome is unlikely to be verified by a confession. There is less opportunity for false negatives

Table 1. Selected Polygraph Validation Studies with Various Methodological Considerations

Reference	Selection of Cases	PJ Ind
Ginton et al. (1982) Journal of Applied Psychology (JAP) 67, 131-137. A method for evaluating the use of the polygraph in real-life situations.	<i>Real life</i> lab study. N=15 final sample. Attempted to remedy some past defects in polygraph studies (p. 132). Examined accuracies with different levels of info. Also attempted to separate the contribution of behavioral cues from the physiological info.	X, 3 had access to different info
Patrick & Iacono (1989) JAP 74, 347-355. Psychopathy, threat, and polygraph test accuracy.	(Lab) Mock theft designed to see if psychopaths were "immune" to detection by polygraph. Tested under conditions of threat.	X, 2 indiv. exams, blind analysis by other examiner
Iacono et al. (1992) JAP 77, 60-64. Use of antianxiety drugs as countermeasures in the detection of guilty knowledge.	(Lab) N=75. Subjects also took APQ <sup>1</sup> (for some) drug study.	X blind
Lieblich et al. (1976) JAP 64, 89-93. Validity of GKT in prisoner's sample.	(Lab study) Used prisoners in jail setting. 1st study to do this with GKT. Subjects listened to stimuli without responding verbally. (matching task)	X
Fornan & McCauley (1986) JAP 71, 691-698.	(Lab) study with undergrads. Author was examiner. Compared CQT <sup>2</sup> to GKT <sup>3</sup> to PCT <sup>4</sup> .	X
Patrick & Iacono (1991) JAP 76, 229-238. Validity of the LQ polygraph test: The problem of sampling bias.	(Field) N=276-402. Considered <i>all</i> cases available for nine largest RCMP detachments from 1980-1984 study. Main issue of test was whether subjects were directly involved.	X blind scoring by 2 independent examiners
Elaad (1990). Detection of guilty knowledge in real-life criminal investigations. JAP, 75, 521-529.	Actual criminal polygraph records randomly sampled. Guilt or innocence of subject was verified by confession of whoever committed crime. Attempted to justify use of confession-verified criterion by comparing demographics of verified and non-verified sample.	X
Kircher and Raskin (1988). Human vs computerized evaluations of polygraph in a lab setting. JAP, 73, 291-302.	(Lab study, w/ local subjects) N=7. Explored hypothesis that statistical approach to polygraph data analysis may improve diagnostic accuracy by eliminating random errors of human judgment. Used a mock crime setting. Good description of physiological measures.	X
Waid and Orne (1980). Individual differences in electrodermal lability and the detection of information and deception. JAP, 65, 1-8.	(Lab study, random assignment) N=28. Electrodermal lability => frequency with which EDRs occur. Frequent responders labeled => labiles. Infrequent responders labeled => stabiles. Hypotheses => guilty stabiles may be more likely to elude detection. Innocent labiles will be more likely to be false positives.	X
Janisse & Bradley (1980). Deception, information, and the pupillary response. Perceptual and Motor Skills, 50, 748-750.	(Lab study, random assignment) N=96. Study designed to investigate the efficacy of the pupillary response in detection of deception, nature of the response, and personality characteristics (Machiavellianism scale). Examiner Judge	X

Amount/Type of Information		Methodology		Physiological Measures				Establishing Ground Truth		Base Rate					
Chart Alone	Preexam Chart	Interview	Succesive R/I	CQT	GKT	Other	SRR	MR/ Cardio	Voice Stress	Other	Confession Verified	Obj. PJ	Prop. CO	Exam ID	Guilty Know.
X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub> Other observed from adjacent room	X				X	X	X			X	.13	No	
			X				X	X	X			X	.50	No	
X			X				X	X	X			X	.59	No	
X			X				X					0	.20	NR probably	
X			X	X	PCT <sup>4</sup>	X	X	X				X	.58	Not clear	
X	X	X	X				X	X	X		X		.58	No	
													.04	No	
													.85	No	
X			X				X			X		X	.49	No	
X	X		X				X	X	X	BP (Blood Pressure)		X	.50	No	
X (exp. 1/2)			X <sub>1</sub> exp. 1	X <sub>2</sub> exp. 2	X						X (exp. 1)	.64 .50	NR probably not		
X			X							Pupillary		X	1.0	1/3 conditions yes-knew what target response	

**Table 1. Selected Polygraph Validation with Various Methodological Considerations (continued)**

Reference	Selection of Cases	Examiner Judge	
		PJ	Ind
Dawson (1980). Physiological detection of responses to questions and answers during countermeasure maneuvers. <i>Psychophysiology</i> , 17, 8-17.	(Lab study) N=24, 2 purposes. (1) test validity of CQT with subjects (actors) who may be able to beat the polygraph test, (2) Separately evaluate physiological responses that occur during two phases, Questions (hearing) and Answers (Verbal responses).		X
Correa and Adams (1981). Validity of the pre-employment polygraph examination and the effect of motivation. <i>Polygraph</i> , 10, 143-155.	(Lab study) N=40. Four conditions 10-lying,motivated; 10-lying, unmotivated; 10-not lying,motivated; and 10-not lying, unmotivated.	X	
Abrams and Weinstein (1974). The validity of the polygraph with retardates. <i>J of Police Science and Administration</i> , 2, 11-14.	(Lab study) N=16. Four different levels of IQ. Examiners could only make judgments on 7/16 subjects, the rest were eliminated from the sample.	X	
Baesan, Chung, and Yang (1948-1949). A lie detector experiment. <i>J of Criminal Law, Criminology, and Police Science</i> , 39, 532-537.	(Class) N=100 experiment that attempted to differentiate between actual guilt of mock crime and simply having knowledge of said crime.	X	
Blum and Osterloh (1968). Polygraph examinations as a means for detecting truth and falsehood in stories presented by police informants. <i>J of Criminal Law, Criminology, and Police Science</i> . 59, 133-137.	(Lab study) N=20. "Good" informants were recruited by police officers. Demand characteristics probably played a significant role.	X	
Honts, Hodes, and Raskin (1985). Effects of physical countermeasures on the physiological detection of deception. <i>JAP</i> , 70, 177-187.	(Lab study) Exp 1, N=52; Exp 2, N=65. Mock theft study in which participants were trained in counter measures. Exp 1 ==> simple behavioral countermeasures. Exp 2 ==> multiple behavioral countermeasures (also included a blood pressure measure which is thought to aid in detecting use of countermeasures).	X	
Honts, Hodes, and Raskin (1985). Effects of physical countermeasures on the physiological detection of deception. <i>JAP</i> , 70, 177-187.	(Lab study) Exp 1, N=52; Exp 2, N=65. Mock theft study in which participants were trained in counter measures. Exp 1 ==> simple behavioral countermeasures. Exp 2 ==> multiple behavioral countermeasures (also included a blood pressure measure which is thought to aid in detecting use of countermeasures).	X, 2 blind ex., 1 original ex.	
Bersh (1969). A validation study of polygraph examiner judgments. <i>JAP</i> , 53, 399-403.	(Field study) N=157. Cases drawn randomly from pool of criminal investigations from 1963-1966. Final.	X	
Barland and Raskin (1976). Validity and reliability of polygraph examinations of criminal suspects. Dissertation study.	Field study. N=102. Requested cases from police, Defense attorneys, or Prosecuting attorneys in Utah and Nevada. Extended findings of Bersh (1969) using criminal suspects tested by Raskin. 92 cases were independent of one another.	X	
Bradley and Ainsworth (1984). Alcohol and the psychophysiological detection of deception. <i>Psychophysiology</i> , 21, 63-71.	(Lab study) N=40. Investigated the effect of intoxication, both during commitment of the crime and interrogation. 2 (sober/intoxicated) $\chi^2$ (guilty/innocent) $\chi^2$ (exam only/crime only/both) $\chi^2$ (order of exams - CQT 1st/GKT 1st) design. BAC=0.12.	X "blind" see p. 66	

**Table 1. Selected Polygraph Validation with Various Methodological Considerations (continued)**

Amount/Type of Information	Methodology	Physiological Measures						Establishing Ground Truth			Base Rate		
		Chart Alone	Preexam	Interview	Succes-	Suc-	MR/ Cardio	Voice Stress	Other	Confession Verified	Obj. PJ	Prop. CO	Exam ID
None cited	X 2 versions - 1 immed. resp./ 1 delayed resp.			X X	X						X	.50	NR probably
X		X		X X	X	X (EKG)					X	.50	yes
X X		probably R/I not reported		X X	X	not specified					X	5 of those left in study	yes
X X		X	POT <sup>7</sup>		X X						X	.50	yes
	X			X X X	X not specified						X	.55	N
X X		X		X X X	X						X	.75	yes--1 blind evaluator
Note: In both experiments, examiner made decision based on pre-exam and polygraph charts. In addition, the same examiner blindly re-evaluated the charts alone 2 weeks later.													
X orig. ex.		X		X X X	X		BP				X	.67	NR probably
X	X (ZOC <sup>5</sup> )	X (GQT <sup>4</sup> )		not reported, (assume GSR)				X 4 members	.44	no overall			
X		X in all but one case		X X X X				X X 5 members	.77	no overall average			
X		X X		X X X					X	.25	NR probably		

to appear in a confession-verified sample. Because innocent people are unlikely to confess following a failed polygraph, such errors are likely to be systematically removed from studies relying on confessions. Thus, the outcome of the polygraph examination will be seen as correct. It is necessary to follow-up as completely as possible *all* cases included in the validation study in order to accurately assess the validity of the polygraph examination and avoid the inflation of accuracy rates due to sampling bias.

A panel judgment is a less frequently used criterion for establishing ground truth (e.g., Bersh, 1969). In this method a panel of judges reviews all the evidence pertaining to a case (except confessions) and issues a judgment of guilty, innocent, or inconclusive. This method does not avoid the problem of sampling bias associated with confession-verified cases when confessions following failed polygraphs are given to the panelists and it is not possible to know whether or not the judgment reached by the panel is accurate. Barland and Raskin (1976) attained polygraph accuracy rates of 99.4% for guilty subjects and 36.5% for innocent subjects (excluding inconclusives) using a panel judgment as the criterion for ground truth. Bersh (1969) measured the validity of polygraph examiner judgments by determining level of agreement with a Judge Advocate General panel. He attained polygraph accuracy rates of 92.9% for guilty subjects and 92.0% for innocent subjects (excluding inconclusives) among cases for which the panel judgment was unanimous and in all likelihood were unanimous because they included confessions derived from failed polygraphs. However, level of accuracy decreased appreciably when panel judgment decreased from unanimous to majority agreement (70.5% and 80.0%, respectively). This study has been criticized because blind chart scoring was not used (Iacono & Patrick, 1988).

Court outcome is a seldom used criterion for establishing ground truth. This is because court outcome may not be a very accurate measure of guilt, due to characteristics of our legal system, such as the necessity for proof beyond reasonable doubt, and the prevalence of plea bargaining. Barland and Raskin (1976) attained accuracy rates of 100% for guilty subjects and 24.6% for innocent subjects (excluding inconclusives) using court outcome as the criterion for ground truth. However, Barland and Raskin were only able to make decisions in 45% of their original

cases, implying that court outcome may not be a very accurate or dependable criterion for ground truth.

### **Individual Differences in Detection of Deception**

Several types of individual differences in susceptibility to detection of deception have been studied. Past research has found that those who scored lower on the Socialization Scale of the California Psychological Inventory, a construct related to psychopathy, were less easily detected as deceptive using a polygraph test (Waid, Orne, & Wilson, 1979). However, more recent research (Patrick & Iacono, 1989) suggests that there are no differences in susceptibility to detection of deception between psychopaths and non-psychopaths. Patrick and Iacono (1989) found that subjects who tried to portray a "stoic facade" during the polygraph test were more likely to be false positives than those who did not. Thus, a substantial proportion of innocent subjects who attempted to hide their fear behind a facade of stoicism, when they were genuinely concerned about the outcome of the polygraph test reacted more strongly to the relevant questions, resulting in an innocent hit rate barely above chance (56%).

Waid and Orne (1980) studied the effects of individual differences in electrodermal lability (the frequency with which electrodermal responses (EDRs) occur) in the detection of deception. They found that stabiles (subjects with relatively fewer EDRs) were less susceptible to detection than were labiles (subjects with relatively more frequent EDRs). Also, labiles were more likely to be a false positive. While Waid and Orne acknowledge that these laboratory results may not generalize to real-life situations, the implication that some subjects may be more likely to be judged DI simply because they are labile needs to be accounted for in any polygraph validation study.

### **Amount of Information Available to Examiner**

In order to assess the validity of the polygraph, it is necessary that judgments be based on the polygraph charts alone. Most studies, both in the laboratory and in the field have "blind examiners" review the polygraph charts and make decisions based solely on the charts. The majority of these studies have found high reliabilities between the original examiners and the blind exam-

iners. Iacono (1991) suggests, however, that this may not provide enough information to establish the validity of the polygraph because of the way in which polygraph tests are conducted. Usually the polygraph examiner stops interviewing subjects when he or she finds DI. Since the polygrapher does not interview any more suspects after someone has been judged DI it is not surprising that polygraph examiners trained in the same manner are able to corroborate each others judgments. This provides all the more reason to follow up cases of subjects who take the polygraph test.

Another issue that is only now beginning to be recognized as a problem is that examiners are not sampled from the population of interest (Barland, 1988). In order for the results of any study to be generalizable to the entire population of polygraph examiners, it is necessary that (as much as possible) the examiners be randomly assigned to subjects. Kircher, Horowitz, and Raskin (1987) found that the decision policy of the examiner accounted for 45% of the variance in mock-crime analog studies. Barland (1988) suggests that a double blind detection study in which neither the subjects nor the examiners are aware that ground truth had been established prior to the polygraph examinations.

Another issue is the extent to which extra-polygraphic cues (e.g., subject's behavior during the pre-polygraph interview) play a role in an examiner's ability to detect deception. Ginton et al. (1982) attempted to separate the contributions of behavioral cues from the physiological information by having independent observers watch the polygraph procedure from behind a one-way mirror. They found that observers were able to detect deception almost as accurately as the polygraphers conducting the examination. This suggests that it is important to determine exactly what behavioral cues the polygrapher attends to during the examination, and how these cues are used in making his or her judgment.

### **Use of Countermeasures**

Lykken (1979) has suggested that it should be relatively easy to defeat a polygraph test. Honts, Hodes, and Raskin (1985) reported that subjects who used a single behavioral countermeasure were not effective in defeating the polygraph (CQT), but subjects who used multiple behavioral

techniques were able to defeat the polygraph test. Furthermore, examiners were unsuccessful in detecting which subjects were using countermeasures. Researchers have also examined the extent to which antianxiety drugs can be used to defeat the polygraph test. Iacono et al. (1992) showed that meprobamate, diazepam, and propanolol were all ineffective in defeating the GKT in a laboratory study.

Bradley and Ainsworth (1984) demonstrated in a laboratory study that subjects who were intoxicated while committing a mock crime were less frequently judged as deceptive in a CQT than those who committed the crime while sober. While these results may not generalize to a real-life polygraph testing situation, the implication is that those who commit crimes while under the influence of alcohol or other drugs may be less susceptible to detection of deception by the polygraph. This implication should be taken into consideration when conducting any validation study of the polygraph.

### **Base Rate Information**

The proportion of guilty subjects in the sample may have an important impact on the accuracy of polygraph judgments. Murphy (1987) argues that the polygraph is often used in situations in which the base rate of deception (or proportion of those guilty) is likely to be very low (as low as 5% in one estimate). If the likelihood that any one person is guilty is very low then the probability is high that many innocent people will be erroneously judged DI. If such a case, the use of the polygraph inappropriate because the polygraph must exhibit extraordinarily high levels of accuracy to avoid showing large number of false positives. It is also vital that the polygraph examiners not have knowledge of the proportion of guilty subjects. If the polygraph examiner knows what portion of subjects in the sample are guilty, he or she is likely to use this information as an aid in making decisions in cases that might otherwise be judged inconclusive (Patrick & Iacono, 1991).

### **Good Validation Study Procedures**

It is important that only well-trained, experienced polygraph examiners are used in a validation study. The training should come from an accredited polygraph training institution. Patrick &

Iacono (1991) used Royal Canadian Mounted Police (RCMP) examiners that had: (1) at least 12 years prior police experience; (2) 5 years' experience in serious crime police investigation; and (3) completed a rigorous polygraph training program. Considering the high quality of the training of military polygraph examiners, satisfying this criterion should not be a difficult task.

It is necessary to sample as many polygraph cases as possible. One problem with past validation efforts is the exclusion of cases on questionable (or not reported) grounds (e.g., Yankee, Powell, & Newland, 1986). Patrick and Iacono (1991) sampled ALL of the cases from the RCMP polygraph files from 1980 to 1984. This a good model to follow in sampling cases. Specific criteria for excluding cases from the sample should be used. When cases are excluded, this should be clearly documented.

Although difficult to do, as many cases as possible should be followed up in order to determine guilt or innocence independently of the polygraph. Patrick and Iacono (1991) were able to verify the guilt or innocence of 75 out of their original 402 cases, independently of the polygraph examination. They accomplished this by collecting as much detailed case information as possible from police detachment files for each of the cases in their sample.

All guidelines followed in classifying polygraph examinees (DI or NDI) should be reported in detail. For example, if numerical scoring is used, the exact procedures used should be reported. All of the examiners in the study should use the same type of chart scoring procedure. A blind chart scoring procedure, in which a second (and possibly a third) group of examiners has access to only the polygraph charts should also be used. As much as possible, the pre-test interview and the construction of questions should be standardized across all examiners in the study.

## **GENERAL PROCEDURES FOR USE OF THE POLYGRAPH IN MILITARY SETTINGS**

### **Some Differences Between Military and Civilian Settings**

Several factors facilitate criminal investigations conducted in the military. Suspects are often more likely to be identified since movement into and out of restricted areas is monitored closely. There is greater opportunity to identify the population of suspects when one knows which individuals were in a given area at a specified time. Presumably all suspects could be polygraphed, which is unlikely in civilian settings. The military code of honor should also motivate witnesses to come forward and tell the truth regarding criminal activities observed. There also may be a greater likelihood that witnesses can be located and will be willing to testify.

Finally, it is less likely that suspects will "disappear" in military settings since they are expected to report for their regular duties. Civilian suspects have more opportunity to flee the scene of the crime and never return.

### **General Use of the Polygraph in the Military**

Detailed procedures for use of the polygraph in the military may be found in Appendices A (Air Force), B (Army), and C (Navy and Marines). For each branch, information is provided in the respective appendix regarding the polygraph's applicability, uses, authorization for approval, authorization forms and criteria, administration procedure, equipment maintenance, reporting method, routing records, and supervision and review.

The present section provides the general format followed when polygraph exams are administered in the military for criminal cases.

An investigation is conducted concerning a crime which has allegedly been committed. At some point, after a fair amount of investigative work has been done, the suspect may be asked to undergo a polygraph exam. The crime must entail some minimal level of consequences (such as potential for death or > 1 year sentence) before a polygraph will be approved. The request must be approved by higher levels before an exam can be administered. That is, not any Joe MP can decide that a polygraph should be given. The approval comes through official channels using official forms.

Once the approval is received, the subject is asked to submit to the polygraph. He or she may refuse with no adverse consequences. (In security clearance cases, the person may be denied the clearance after refusing. This is not considered to be an adverse consequence as long as a position of equal pay and status is maintained or provided.) A point is made, however, that the investigation may continue even if the subject refuses to take the exam. Any person who is the subject of an investigation may request an exculpatory polygraph exam and their request must be honored.

Before the polygraph examination, the examinee is advised of his or her rights. For example, they are told they may consult with legal counsel or they may refuse to undergo the polygraph with no adverse consequences. The Army and Air Force specify that the general polygraph procedure must be explained and the examination questions must be reviewed before the actual testing begins.

For administering the examination, there are guidelines about what types of questions are prohibited (at least in the Army and Air Force). There are fairly standard forms (across the military) which are filled out during the exam describing the questions asked, the responses, and a signed release form that the examinee signs and which says that he or she understands their rights with regard to the polygraph exam. A summary of the timing and procedures followed during the exam, a question by question description of the scoring, and an overall judgment about deception are also included in the report.

Only certified examiners may conduct the exam in the Army and Air Force. The Navy/Marines did not specify this in their procedures, but this appears to be true for them as well. After the exam, the examiner scores the polygraph and then the results must be sent to a central clearinghouse for review, at least in the Army and Air Force. Again, the Navy/Marines didn't specify this but it appears to be done by them as well. In almost all cases it appears polygraph results are re-released to the investigators after the chart(s) have been through quality control but the regulations don't actually specify this.

The polygraph regulations are silent regarding attempts by examiners to interrogate or extract a confession from the examinee. The Army does say that the polygraph can't be used as a psychological prop in the conduct of interrogations.

### **Polygraph Procedures: Specific Comparisons Between Branches**

#### **Applicability**

The polygraph's applicability is about the same for all three branches. Each branch's procedures are applicable to military personnel within that branch. However, the Air Force's applicability appears to be broader, covering all U.S. citizens and foreign nationals.

#### **Uses**

The Army and Air Force procedures specify that some types of questions are prohibited, e.g., questions about religious beliefs, political views, etc. The Navy/Marines' procedures do not specify any such prohibitions.

The Army procedures specify that exculpatory examinations may be requested by the suspect(s), even if the case may be disposed of under a penalty that is lower than that required for a substantive polygraph exam to be approved (i.e., death or greater than one year imprisonment). The Navy/Marines and Air Force procedures do not make a point of this (although their regulations do not exclude it). The Navy/Marines' procedures make a point that there must be no evidence of implied pressure or coercion when an exculpatory examination is requested; the Army and Air Force procedures do not make that specific point.

The Navy/Marines procedures are not as specific as the Army or Air Force procedures, in general, about *who* may be examined (e.g., military members, civilians, foreign nationals). The Army and Air Force procedures make the disclaimer that use of the polygraph must also obey the laws of the foreign country in which it is to be used; the Navy/Marines' procedures do not. The Air Force procedures are more detailed than the procedures from the other two branches about when the polygraph may be used. The Air Force procedures do an especially thorough job of specifying

ing the circumstances under which polygraphs may be used in criminal investigations - i.e., when all other avenues of investigation have been tried, when there is good reason to believe that the person is involved in the crime, etc.

When the polygraph is used to investigate credible derogatory information about someone or some organization, the Navy/Marines' procedures specify that the polygraph should be used only if no other means of resolution is available. The Army procedures do not explicitly state this. The Air Force procedures include such a requirement for the conduct of counterintelligence operations but do not mention investigating derogatory information specifically.

The Army procedures specify that the polygraph may be used to examine foreign nationals whose background cannot be verified but who require access to classified data. The Navy/Marines and Air Force procedures do not mention this situation.

The Air Force procedures specify that persons who have already been court-martialled for the offense in question may not be given a polygraph examination (except in very rare circumstances). The Army and Navy/Marines' procedures do not say anything about this situation.

#### **Authorization (for Approval)**

The Air Force delegates the authority to "offer" (i.e., "approve") a polygraph to persons lower in the chain of command (detachment commanders) than do the other two branches.

#### **Authorization (Forms and Criteria)**

The kind of information which must be included in the request for a polygraph exam is basically the same across branches. One difference appears to be that the Army requires more proof, documentation, explanation that the investigation prior to the polygraph request has been as thorough as possible. The Navy/Marines and Air Force require a copy of all the investigation materials and an explanation of why the exam is deemed essential.

The Air Force procedures specifically mention that a Medical Records Check must be made.

The Army and Navy/Marines procedures do not.

The Air Force procedures mention the situation in which Air Force personnel review results of polygraph examinations conducted by civilian examiners. The procedures recommend that Air Force commanders do not consider results of such examinations unless the qualifications of the examiner, charts, questions, and techniques used have been reviewed by a DOD certified examiner who says the exam meets Air Force standards. Neither the Army nor the Navy/Marines procedures mention this situation.

The Air Force procedure mention that a Defense Central Index of Investigations Check will be conducted on the subject, victim, or proposed examinees of each polygraph request prior to approval of the request. The procedures to the other branches do not mention this.

The Air Force procedures describe what to do (as far as record-keeping) if the subject declines to take the polygraph exam *after* the request was approved. The procedures for the other branches do not.

The Marines procedures mention using telephone requests for polygraph approval, especially in emergencies, but the Navy procedures do not mention using AUTOVON. Both the Army and Air Force procedures mention using the telephone to make requests in an emergency.

### **Administration Procedure**

The Navy/Marines' procedures provide relatively little detail in this section, the Army procedures go into quite a bit of detail, and the Air Force procedures go into excruciating detail, including information about the conduct of the *investigator* prior to the polygraph examination.

The Army and Air Force procedures discuss requirements for the polygraph examiner (e.g., certification, supervised internship, no special titles or clothing) and specify that the polygraph examiner is to receive all investigation data on the examinee prior to the examination. They also specify that the polygraph is not to be used as a prop for interrogation, that the examinee is not to be subjected to prolonged interrogation immediately prior to the exam (notice but say nothing

about after the exam), that the polygraph instrument and procedures will be explained, that all exam questions will be reviewed prior to testing, that female examinees will be examined in the presence of a female witness, or that opinions on truth or deception can be formed only when each relevant question has been asked on more than one chart. The Navy/Marines procedures do not discuss any of these issues.

As mentioned above, the Air Force procedures provide guidelines for the investigator. The Air Force procedures also discuss how to fill out the necessary forms and how to mark the polygraph chart appropriately. Procedures to be followed during the pre-examination, and post-testing phases are discussed as well. Neither the Navy/Marines' procedures nor the Army procedures go into such detail.

***Equipment Maintenance.*** The Navy/Marines' procedures are not as specific regarding equipment maintenance as the Army or the Air Force. The Navy/Marines' procedures merely say that the polygraph instrument should be calibrated prior to the examination (they do not specify calibration prior to *each* exam or how long before the exam the calibration must be done). A calibration chart is included as part of the record of every exam. The only other thing the Navy/Marines' procedures say is that CMC should be notified for major maintenance/repairs. The Army procedures go into detail about how to calibrate the instrument, that calibration must occur within the 24 hours prior to each exam, securing machines between exams, and the standards which must be met by any new equipment. The Air Force procedures specify how to procure the equipment, that the polygraph must be calibrated within the 24 hours preceding each exam, and that minor repairs should be performed by the examiner while major repairs should be referred to HQ AFOSI/IVSP.

***Reporting Method.*** The Navy/Marines' procedures specify the name of the Report which should be completed. The Army procedures discuss the different forms that must be filled out, the information which goes on each form, and who receives copies. The Air Force procedures go into even more detail than the Army procedures about which forms must be filled out and how they should be completed.

**Routing Records.** The Army and Air Force procedures discuss where all the records are sent and how they are filed. The Navy/Marines' procedures discuss who can get access to the results and mention sending the records to the central office of NIS.

The Air Force procedures specify where a declination of a substantive examination may be recorded. It can be recorded in the inner page of a report which is not routinely released outside of AFOSI. This appears to contradict the declaration in Section II of the Air Force regulations that says refusals will not be placed anywhere in personnel files.

The Air Force procedures also describe a Management Effectiveness Profiling System (MEPS) which sounds like an accounting-for-resources-expended system. The procedures from the two other branches do not mention this kind of system, although the Army procedures do talk about depositing records at a central clearinghouse.

**Supervision and Review.** The Navy/Marines' procedures say that a supervisory official who is a certified polygrapher will exercise professional and technical supervision. They do not say anything about quality control procedures. The Army procedures talk about how each set of charts is reviewed by supervisors. The reviewers may directly contact the original examiner, and if the two do not interpret the charts in the same way, the supervisor may require the exam to be repeated. For polygraphs found to be "inconclusive" or "no opinion," a repeat exam may be administered without having to request approval again (it's unclear if this applies to the original judgment, to the reviewer's judgment, or to both). The Air Force procedures say that each case is reviewed by a quality control supervisor who prepares a written critique of each examination and passes that information along to the original examiner and HQ. Based on this critique, HQ decides whether the exam should be repeated or not. The Air Force procedures also specify that for counterintelligence security polygraph exams, a second examiner must be *present* during the examination and must conduct an immediate on-scene numerical evaluation and quality control review after the first examiner conducts the exam and scores the chart.

### **Computer File Maintenance of Polygraph Data**

Each military branch stores its own polygraph data in a central repository. The Marine polygraph data is recorded in a logbook but the actual files are stored with the Navy database. Relevant details from the criminal investigative files and outcomes of the polygraph are transferred to and stored on computers.

Each branch stores different information from the case files and polygraph charts. Table 2 provides the types of information available on the computer polygraph files by military branch.

All military branches store the following information in their computer files:

1. Purpose of exam (i.e., specific incident for criminal investigation or security screening)
2. Criminal code category
3. Polygraph Initiation date and exam date
4. Polygraph findings (DI, NDI, Inconclusive, No Opinion)
5. Confirmation Status or Confirmation Method

Additionally, each branch uses a different computer software database package to store and retrieve the polygraph information. The Air Force provided files using the CONDOR database, the Navy provided files using the dBase database, and the Army provided files in ASCII format. The version of CONDOR used by the Air Force is not available in the public domain and special arrangements had to be made to retrieve Air Force data for this study. The Marine data is stored by the Navy, although the Marines keep a written log of Marine cases.

The lack of a centralized database across all branches significantly hampered the present study's data collection efforts.

Table 2

## Information Available on the Computer Polygraph Files by Branch\*

Information	AF	Army	Marines	Navy
1. Purpose (i.e., specific incident)	Yes	Yes	Yes	Yes
2. Criminal Code				
A. Category	Yes	Yes	Yes	Yes
B. Esculpatory	Yes	No	Yes	Yes
3. Input Data				
A. Facts of Case	No	No	No	No
B. Demographic Characteristics				
(01) Prior exam	No	No	Yes	No
(02) Age	No	No	No	Yes
(03) Sex	No	No	No	Yes
(04) USA/Foreign	Yes	Yes	No	Yes
(05) Race	No	No	No	Yes
(06) Examinee Type (i.e., Subject, Witness, Victim)	Yes	No	No	Yes
(07) Education	No	No	No	Yes
(08) Service Time	No	No	No	Yes
(09) Military/Civilian	Yes	Yes	No	Yes
(10) Rank	Yes	Yes	No	Yes
(11) Grade	No	No	No	Yes
(12) Requester	Yes	No	No	Yes
(13) Clearance	No	No	No	Yes
4. Process Data				
A. Steps Completed	No	No	No	No
B. Polygraph Dates/Times				
(01) Initiation Date	Yes	Yes	Yes	Yes
(02) Approval Date	Yes	No	Yes	Yes
(03) Exam Date	Yes	Yes	Yes	Yes
(04) Length of Exam	No	Yes	No	Yes
(05) Quality Control				

Table 2 continued

## Information Available on the Computer Polygraph Files by Branch\*

Information	AF	Army	Marines	Navy
Review Date	Yes	No	Yes	Yes
(06) Rec'd at HQ Date	Yes	No	Yes	Yes
(01) Technique Used	No	Yes	Yes	Yes
(02) Examiner Code	Yes	Yes	Yes	No
(03) No. of Techniques Used	No	No	Yes	No
(04) Number of Charts	No	No	No	Yes
(05) Area Administered	Yes	No	No	Yes
(06) Requesting Office	Yes	No	No	Yes
(07) Travel Time	No	No	No	Yes
5. Outcome Data				
A. Polygraph Findings	Yes	Yes	Yes	Yes
B. Additional Leads as a result of polygraph	Yes	No	No	No
C. Quality Control Outcome	No	No	Yes	Yes
D. Urinalysis	Yes	No	No	No
E. Ground Truth Criteria				
(01) Confirmation Status	Yes	Yes	No	Yes
(02) Method of Confirmation				
a. Pre-test Confession	Yes	No	Yes	Yes
b. Pre-test Admission	No	No	Yes	No
c. Post-test Confession	Yes	No	Yes	Yes
d. Post-test Admission	No	No	Yes	No
F. Investigative Outcome	No	No	No	No

\* Note: The polygraph data from the Marines were obtained from photocopies of the written log books.

Since different types of data are collected and then stored in different databases using different software, any polygraph research conducted across military branches will pose significant obstacles to future researchers.

It is a strong recommendation of this report to establish a centralized database using standard formats for data retrieval and storage. Likewise, standard types of information should be collected from all branches.

### **Analyses of Military Computer Files**

In order to establish base rates of polygraph cases that have been independently confirmed or unconfirmed by confession, the computer files provided to PDRI by the various branches of military service were analyzed. Although the information present in the computer files for each branch varied, as had been previously mentioned, all branches provided information on the results of the polygraph exam (i.e., Deception Indicated, No Deception Indicated, Inconclusive, and No Opinion) and confirmation status (i.e., whether or not the polygraph results were Confirmed or Unconfirmed). After considerable work deciphering the various databases (the various branches differed in the type of operating system, the type of database package employed, the coding and formatting of information contained in the data set, the variable field lengths, and the manner by which case data was aggregated), the data had been analyzed. PDRI received Air Force computer files for the period 1986-1989. The Air Force files provided information on 4838 cases in that time period. The Army computer files contained 19,804 cases for the period 1983-1989; the Marines provided 837 cases from 1987-1989, and the Navy provided 9839 cases for 1980-1987.

Each military branch has its own coding system and method for organizing crimes into categories. As part of the process of creating a common taxonomy across branches, Appendix D provides a detailed comparison of the criminal investigative categories and corresponding codes used by the various military services. The Army has the most detailed coding system for classifying crimes.

### **Frequency and Cross-Tabulation of Crime Category by Year by Military Branch.**

Each branch's polygraph database was analyzed separately. The first series of analyses generated frequency counts for each crime category, and then cross-tabulated across different years. The analyses of the frequencies of the various types of crimes by year may be found in Table 3 for the Air Force, Table 4 for the Army, Table 5 for the Marines, and Table 6 for the Navy.

A detailed breakdown of the crosstabulation counts for each crime category by year may be found in Appendix E for the Air Force, Appendix F for the Army, Appendix G for the Marines, and Appendix H for the Navy. Please note the frequency counts are based on computer data received by PDRI at the time of its request to each military ranch in 1989. Data for 1989 and in some cases the previous years were incomplete since a number of cases were still open at the time of the request. Also the interval between when a case was closed and its entry into the computer database varied across branches.

The criminal categories with the highest percentages were "sex offenses" (40.09%) for the Air Force (Table 3); "Crimes against property (51.57%) for the Army (Table 4); "Larceny of personal property" (51.72%) for the Marines (Table 5), and "Polygraph Examination (49.78%) for the Navy (Table 6). The "Polygraph Examination" category for the Navy included drug-related cases involving urinalysis exams and requests for the polygraph for exculpatory purposes.

### **Polygraph Results by Military Branch**

To obtain information regarding base rates and hit rates, the polygraph results were compared to confirmation status for each military branch for each year. The polygraph results by confirmation tables for the Air Force for the years 1986-1989 is found in Appendix I. Analogous tables for the Army (1983-1989) may be found in Appendix J; for the Marines (1987-1989) in Appendix K; and for the Navy (1980-1987) in Appendix L. Tables 7A-7G summarizes the polygraph results by confirmation status by year across all branches. In Tables 7A-7G, the numbers in

Table 3

Frequency of Crime by Criminal Category  
 (Air Force Data: 1987-1988)

Code	Category Description	Frequency	%
3	Compromise of Test material	19	1.76
6	Death Investigations	66	6.11
7	Sex Offenses	433	40.09
8	Assault	62	5.74
9	Impersonation	1	.09
10	Fraudulent Enlistment, Appointment,Discharge or Separation	8	.74
12	Forgery	30	2.78
13	Larceny	42	3.89
14	Theft of Govt.Property	63	5.83
15	Robbery	3	.28
16	Housebreaking	31	2.87
17	Drug Abuse	200	18.52
19	Blackmarket Activities	4	.37
20	Property Destruction	99	9.17
22	Intimidation	19	1.76
<b>Totals</b>		<b>1080</b>	<b>100</b>

Table 4  
Frequency of Crime by Criminal Category  
(Army Data: 1983-1989)

Code	Category Description	Freq.	%
5	CRIMES AGAINST PERSONS (Subtotal)	6523	32.94
5A	Adultery.....	105	0.53
5B	Bigamy .....	12	0.06
5C	Assaults (Other Than Sexual) .....	424	2.14
5D	Child Abuse.....	121	0.61
5E	Civil Rights .....	0	0.00
5F	Extortion.....	29	0.15
5G	Forgery .....	662	3.34
5H	Murder, Manslaughter, Homicide, Acc. Death .....	312	1.58
5I	Omitted.....	0	0.00
5J	Impersonations .....	4	0.02
5K	Kidnapping.....	47	0.24
5L	Drugs .....	3990	20.15
5M	Perjury .....	108	0.55
5N	Robbery .....	360	1.82
5P	Suicide.....	1	0.01
5Q	Traffic.....	19	0.10
5R	Weapons .....	7	0.04
5S	Escape.....	2	0.01
5T	Communications Incidents .....	160	0.81
5U	Harassing Communications.....	4	0.02
5V	.....	7	0.04
5W	Unauthorized Absences .....	5	0.03
5X	Other Offenses Against Persons.....	51	0.26
5Y	Conduct Unbecoming a Member of Military Serv.....	90	0.45
5Z	.....	3	0.02
6	CRIMES - SEX (Subtotal)	2620	13.23
6A	Child Molestation.....	384	1.94
6B	Homosexuality .....	38	0.19
6C	Indecent Acts.....	667	3.37
6D	Pornography .....	2	0.01
6E	Rape, Carnal Knowledge.....	1064	5.37
6F	Sodomy.....	406	2.05
6G	Bestiality.....	0	0.00
6H	Exhibitionism .....	8	0.04
6J	Incest .....	4	0.02

Table 4 continued

**Frequency of Crime by Criminal Category  
(Army Data: 1983-1989)**

Code	Category Description	Freq.	%
6K	Obscene Communications.....	29	0.15
6M	.....	1	0.01
6N	Criminal Abortion .....	0	0.00
6X	Other Sex Crimes .....	17	0.09
7	<b>CRIMES AGAINST PROPERTY (Subtotal)</b>	<b>10214</b>	<b>51.57</b>
7A	Arson .....	169	0.85
7B	Blackmarket .....	131	0.66
7C	Burglary, Housebreaking, Unlawful Entry .....	2049	10.35
7D	Counterfeiting.....	12	0.06
7E	Smuggling .....	8	0.04
7F	Larceny, Government Property .....	2877	14.53
7G	Larceny, Private Property/ Funds.....	4266	21.54
7H	Postal Violations .....	221	1.12
7J	Wrongful Obstruction .....	442	2.23
7K	Receiving Stolen Property.....	24	0.12
7L	Failure to Pay Just Debt .....	2	0.01
7X	Other Crimes Against Property .....	13	0.07
8	<b>CRIMES - FRAUD (Subtotal)</b>	<b>433</b>	<b>2.19</b>
8A	Bribery.....	154	0.78
8B	Conflict of Interest .....	10	0.05
8C	Dependency Assistance.....	1	0.01
8D	Personnel Action .....	50	0.25
8E	Non-Appropriated Funds.....	3	0.02
8F	Pay and Allowances .....	106	0.54
8G	Procurement/Salvage/PDO.....	22	0.11
8H	Fraud/Pol items .....	1	0.01
8J	.....	1	0.01
8K	.....	2	0.01
8X	Other Crimes Fraud .....	83	0.42
	<b>Nonexistent Codes/ Coding Errors (Subtotal)</b>	<b>14</b>	<b>0.07</b>
	<b>Total</b>	<b>19804</b>	<b>100.0</b>

Table 5  
Frequency of Crime by Criminal Category  
(Marine Data: 1987 - 1989)

Crime	Category Number		Total Frequency	%
	USMC	PDRII		
Larceny of personal property	6T	0907	435	51.72
Drugs - Narcotics	7N	0710	193	22.95
Larceny of government property	6S	0906	25	2.97
Wrongful property destruction	6U	0912	31	3.69
Assault	7G	0802	31	3.69
Fraud crimes - other (most involve false swearing)	4X	0608	31	3.69
Fraudulent enlistment	4X	0609	1	.12
Forgery	7F	0611	6	.71
Burglary	6N	0902	10	1.79
Sodomy	8G	1010	6	.71
Polygraph examination	9B			
- Screening exams	9B	1904	13	1.55
- Adultery	9B	0801	8	.95
- Fraternization	9B	1904	3	.36
- False swearing	9B	1904	2	.24
- Article 92, sleeping on post	9B	1904	1	.12
- Misconduct	9B	0715	1	.12
- other/unknown/unspecified	9B	1904	3	.36
Obscene phone calls	8X	1014	7	.83
Other sex crimes	8X	1016	4	.48
Sexual assault (indecent assault)	8D	1006	6	.71
Rape	8F	1008	3	.36
Child abuse	7L	0807	4	.48
Child sex abuse	8B	1003	2	.24
Negligent homicide	7H	0813	3	.36
Bribery	4H	0601	2	.24
Robbery	7R	0915	1	.12
Postal violations	6L	0717	1	.12
Perjury	7P	0718	1	.12
Arson	6A	0901	1	.12
Communications incidents (crimes against persons)	8X	0808	1	.12
Other crimes against persons	7X	0821	1	.12
Unspecified crimes			4	.48
<b>Totals</b>			<b>841</b>	<b>100.00</b>

Table 6

## Frequency of Crime by Criminal Category (Navy Data: 1980-1987)

Code	Type of Criminal Case	Total Frequency	%
1.	SPECIAL INQUIRIES	0	0.00
1K	Visa Applicant	0	0.00
1L	Local Security Inquiry	0	0.00
1M	Foreign National Marriage	0	0.00
2.	PERSONNEL INQUIRIES	0	0.00
2A	Agent Applicant	17	.19
2B	Internal Personnel Inquiry		
3.	INTERNAL SECURITY	5	.05
3A	Order and Discipline	183	2.00
3C	Espionage	41	.45
3D	Contact Reports	1	.01
3E	Continuing Interest	0	0.00
3F	Sabotage	1	.01
3G	Information Request	7	.08
3X	Special Inquiry		
4.	CRIMINAL-FRAUD	8	.09
4C	Conflict of Interest	84	.92
4E	Pay and Allowance	15	.16
4F	Personnel Action	27	.29
4G	Procurement	21	.23
4H	Bribery	0	0.00
4I	Dependency Assistance	89	.97
4K	Forgery	0	0.00
4S	Investigative Survey	9	.10
4T	Unauthorized Services	23	.25
4X	Special Inquiry		
5.	SECURITY	145	1.58
5B	Loss of Classified Matter	0	0.00
5C	Ci Port Brief	28	.31
5D	Unauthorized Disclosure	1	.01
5E	Leakage		

Table 6 (continued)

## Frequency of Crime by Criminal Category (Navy Data: 1980-1987)

Code	Type of Criminal Case	Total Frequency	%
5F	Compromise	86	.94
5G	Threat Assessment General	0	0.00
5M	OPSEC Support	3	.03
5S	CI Studies	0	0.00
5T	Terrorist	11	.12
5X	Special Inquiry	35	.38
<b>6. CRIMINAL-PROPERTY</b>			
6A	Arson	97	1.06
6C	Blackmarket	1	.01
6G	Counterfeiting	22	.24
6L	Postal	74	.81
6M	Customs	3	.03
6N	Burglary	348	3.80
6R	Larceny-Ordnance	98	1.07
6S	Larceny-Government	627	6.85
6T	Larceny-Personal	479	5.23
6U	Wrongful Destruction	201	2.20
6V	Larceny-Non Govt. Vehicle	42	.46
6X	Special Inquiry	11	.12
<b>7. CRIMINAL-PERSON</b>			
7A		2	.02
7B	Bomb Threat	51	.56
7D	Criminal Inquiry	0	0.00
7E	Extortion	32	.35
7F		4	.04
7G	Assault	63	.69
7H	Death	62	.68
7K	Kidnapping	3	.03
7L	Child Abuse	43	.47
7N	Narcotics *	355	3.88
7P	Perjury	17	.19
7R	Robbery	35	.38
7T	Traffic Accident	0	0.00
7X	Special Inquiry	30	.33

Table 6 (continued)

## Frequency of Crime by Criminal Category (Navy Data: 1980-1987)

Code	Type of Criminal Case	Total Frequency	%
8.	CRIMINAL-SEX		
8B	Sex Abuse - Child	160	1.75
8D	Indecent Assault	42	.46
8F	Rape	104	1.14
8G	Sodomy	259	2.83
8P		1	.01
8X	Special Inquiry	31	.34
9.	SPECIAL ACTIVITIES		
9A	Protective Services	1	.01
9B	Polygraph Examination **	4557	49.78
9C	Technical Inspection	0	0.00
9D	Technical Survey	0	0.00
9E	TSCM Briefing	0	0.00
9F	Defensive Briefing	0	0.00
9G	Forensic Examination	0	0.00
9M	Technology Transfer CI Briefing	1	.01
9P	Polygraph (CSP)	1	.01
9S	The Collectors	0	0.00
9T	C/T Support	0	0.00
9X	Special Inquiry	0	0.00
IN	(Initial)	202	42.21
MA		1	.01
OG	(Ongoing)	75	.82
OP		11	.12
TF	(Transfer)	10	.11
TR	(Termination)	158	1.73
<b>Totals</b>		<b>9154</b>	<b>100.00</b>

\* not urinalysis

\*\* request for an esculpatory when case closed; also includes all urinalysis

parenthesis represent the percent of individuals in the specific military branch that fall in the particular category. For example, in Table 7A, 57.7% of all the individuals tested in the Army have a DI-Confirmed outcome. In general, about 75% of the Army cases are confirmed by confession whereas about 35-40% of the cases of the other branches are confirmed.

### **Selection of Case Files**

#### **Creating a Taxonomic Structure of Criminal Investigative Cases**

The taxonomic structure which served as a basis for sampling and selecting polygraph case files was, by necessity, derived from the computer files supplied to PDRI by the various branches of military service. Since the present investigators did not have access to the actual case files, the selection procedure was based on the information found in the computer files. To ensure consistency in selecting cases across branches, the taxonomic variables were based on information available from all four branches (i.e., a "Yes" response for each of the four branches in Table 3). By examining the data in Table 3, it was clear that the initial taxonomic structure would include the following four variables:

1. Purpose (i.e., specific-incident criminal case)
2. Criminal Code Category
3. Polygraph Finding (i.e., "DI", "NDI", "Inconclusive") and
4. Confirmation status or confirmation method

#### **Initial Selection of Case Files to Review**

The initial selection of cases from each branch was designed to obtain a variety of cases across different criminal categories. Cases were also selected to reflect different polygraph outcomes. Initially, each branch was requested to provide 25 cases. The case numbers were selected by searching through the computer database.

Table 7A

## Polygraph Result by Confirmation Status Across All Branches - 1983\*

	Confirmed		Unconfirmed		Result Subtotal	
	Army	Navy	Army	Navy	Army	Navy
Deception Indicated	1829 (57.7)	604 (33.6)	453 (14.3)	477 (26.5)	2282 (72.0)	1081 (60.1)
No Deception BIndicated	468 (14.8)	15 (0.8)	421 (13.3)	686 (38.2)	889 (28.0)	701 (39.0)
Inconclusive/ No Opinion	0 (0.0)	2 (0.1)	0 (0.0)	14 (0.8)	0 (0.0)	16 (0.9)
Confirmation Status Subtotal	2297 (72.4)	621 (34.5)	874 (27.6)	1177 (65.5)	3171 (100.0)	1798 (100.0)

\* The data reported for the Navy actually covers 1980 - 1983

Table 7B

## Polygraph Result by Confirmation Status Across All Branches - 1984

	Confirmed		Unconfirmed		Result Subtotal	
	Army	Navy	Army	Navy	Army	Navy
	-----	-----	-----	-----	-----	-----
Deception Indicated	1913 (63.6)	676 (42.1)	363 (12.1)	365 (22.7)	2276 (75.6)	1041 (64.8)
No Deception Indicated	282 (9.4)	12 (0.8)	451 (15.0)	494 (30.7)	733 (24.2)	506 (31.5)
Inconclusive/ No Opinion	0 (0.0)	2 (0.1)	0 (0.0)	58 (3.6)	0 (0.0)	60 (3.7)
Confirmation Status Subtotal	2195 (73.0)	690 (42.9)	814 (27.0)	917 (57.1)	3009 (100.0)	1607 (100.0)

Table 7C

## Polygraph Result by Confirmation Status Across All Branches - 1985

	Confirmed		Unconfirmed		Result Subtotal	
	Army	Navy	Army	Navy	Army	Navy
Deception Indicated	2082 (68.3)	973 (31.9)	332 (10.9)	393 (19.2)	2414 (79.2)	1366 (66.7)
No Deception Indicated	186 (6.1)	13 (0.6)	450 (14.8)	588 (28.7)	636 (20.9)	601 (29.4)
Inconclusive/ No Opinion	0 (0.0)	4 (0.2)	0 (0.0)	76 (3.7)	0 (0.0)	80 (3.9)
Confirmation Status Subtotal	2268 (74.4)	990 (48.4)	782 (25.6)	1057 (51.6)	3050 (100.0)	2047 (100.0)

Table 7D

## Polygraph Result by Confirmation Status Across All Branches - 1986

	Confirmed			Unconfirmed			Result Subtotal		
	AirFor	Army	Navy	AirFor	Army	Navy	AirFor	Army	Navy
Deception Indicated	151 (36.3)	1913 (70.5)	798 (40.0)	56 (13.5)	247 (9.1)	482 (24.2)	207 (49.8)	2160 (79.6)	1280 (64.2)
No Deception Indicated	15 (3.6)	173 (6.4)	19 (1.0)	185 (44.5)	382 (14.1)	583 (29.2)	200 (48.1)	555 (20.4)	602 (30.2)
Inconclusive/ No Opinion	0 (0.0)	0 (0.0)	1 (0.1)	9 (2.2)	0 (0.0)	0 (5.6)	112 (2.2)	9 (2.2)	0 (0.0)
Confirmation Status Subtotal	166 (39.9)	2086 (76.8)	818 (41.0)	250 (60.1)	629 (23.2)	1177 (59.0)	416 (100.0)	2715 (100.0)	1995 (100.0)

Table 7E

## Polygraph Result by Confirmation Status Across All Branches - 1987

	Confirmed				Unconfirmed				Result Subtotal			
	AirFor	Army	Marine	Nav	AirFor	Army	Marine	Navy	AirFor	Army	Marine	Navy
Deception Indicated	490 (34.6)	1586 (64.7)	96 (29.3)	469 (32.9)	216 (15.3)	290 (11.8)	143 (43.6)	364 (25.5)	706 (49.9)	1879 (76.6)	239 (72.9)	833 (58.4)
No Deception Indicated	36 (2.5)	275 (11.2)	1 (.03)	2 (.01)	600 (42.4)	299 (12.2)	68 (20.7)	488 (34.2)	636 (45.0)	574 (23.4)	69 (21.0)	490 (34.4)
Inconclusive/ No Opinion	2 (0.1)	0 (0.0)	1 (0.3)	0 (0.0)	71 (5.0)	0 (0.0)	19 (5.8)	103 (7.2)	73 (5.2)	0 (0.0)	20 (6.1)	103 (7.2)
Confirmation Status Subtotal	528 (37.3)	1864 (76.0)	98 (29.9)	471 (33.0)	887 (62.7)	589 (24.0)	230 (70.1)	955 (67.0)	1415 (100.0)	2453 (100.0)	328 (100.0)	1426 (100.0)

Table 7F

## Polygraph Result by Confirmation Status Across All Branches - 1988

	Confirmed			Unconfirmed			Result Subtotal		
	AirFor	Army	Navy	AirFor	Army	Navy	AirFor	Army	Navy
AirFor	414 (30.3)	1461 (75.5)	102 (35.5)	205 (15.0)	153 (7.4)	104 (36.2)	619 (45.3)	1614 (83.5)	206 (71.8)
Deception Indicated	46 (3.4)	158 (8.2)	0 (0.0)	639 (46.7)	162 (8.4)	52 (18.1)	685 (50.1)	320 (16.6)	52 (18.1)
No Deception Indicated	3 (0.2)	0 (0.0)	2 (0.7)	61 (4.5)	0 (0.0)	27 (9.4)	64 (4.7)	0 (0.0)	29 (10.1)
Inconclusive/ No Opinion									
Confirmation Status Subtotal	463 (33.9)	1619 (83.7)	104 (36.2)	905 (66.2)	315 (16.3)	183 (63.8)	1368 (100.0)	1934 (100.0)	287 (100.0)

Table 7G

## Polygraph Result by Confirmation Status Across All Branches - 1989

	Confirmed			Unconfirmed			Result Subtotal		
	AirFor	Army	Navy	AirFor	Army	Navy	AirFor	Army	Navy
Deception Indicated	57 (31.3)	72 (100.0)	7 (63.6)	32 (17.6)	0 (0.0)	4 (36.4)	89 (48.9)	72 (100.0)	11 (100.0)
No Deception Indicated	14 (7.7)	0 (0.0)	0 (0.0)	78 (42.9)	0 (0.0)	0 (0.0)	92 (50.6)	0 (0.0)	0 (0.0)
Inconclusive/ No Opinion	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	0 (0.0)	0 (0.0)	1 (0.6)	0 (0.0)	0 (0.0)
Confirmation Status Subtotal	71 (39.0)	72 (100.0)	7 (63.6)	111 (61.0)	0 (0.0)	4 (36.4)	182 (100.0)	72 (100.0)	11 (100.0)

After reviewing the initial set of cases, it was determined that cases involving "drugs" were least informative. Investigative files for these cases contained the least information and the polygraph exams were often conducted for exculpatory purposes. There was little basis for independently establishing ground truth with the drug cases except perhaps through comparing polygraph results with urinalysis findings which typically were also included in the case file. Therefore, the decision was made to exclude drug cases from the panel study.

### **Final Selection of Case Files to Review**

Case selection soon focused on multiple-suspect cases since it became apparent that multiple suspect files would be very suitable for presentation to panel members. Since the large majority of single-subject cases were DI-confirmed, multiple suspect cases were believed to be more "balanced" since some subjects would be innocent and others guilty and some subjects polygraphed would be DI and other suspects would be NDI. Most multiple suspect cases were found in the rape, larceny, murder, and child abuse criminal categories.

The criteria, requirements and rationale for case selection are provided below.

#### **1. Year Case was Initiated**

Minimum Requirement: 1985 - 1988 (inclusive)

Preferred Requirement: 1986 - 1987 (inclusive)

Rationale: We can identify cases from our computer files for all service branches during this time frame. The time frame is far enough in the past to allow for files to be complete but not so far past that retrieval becomes a problem. If we request files from more recent years we will have a greater likelihood that the criminal investigation is still on-going and the requested file will not be complete or still "in the field." We may have difficulty ascertaining the ground truth of "in the field" (i.e., no independent or alternate source of verification).

2. Multiple Suspect File

Minimum Requirement: At least two suspects/witnesses were administered the polygraph exam.

Preferred Requirement: More than two suspects/witnesses were administered the polygraph exam.

Rationale: One suspect case files are almost solely DI-Confirmed cases, thus the base rate will be too high. Our panel needs to be presented multiple suspect cases which offer more outcome possibilities.

3. Polygraph Exam Result Mix

Minimum Requirement: One DI suspect and one NDI suspect, *or* two suspects.

Preferred Requirement: More than two suspects with at least one DI suspect and at least one NDI suspect, *or* at least two NDI suspects.

Rationale: We currently have a sufficient number of multiple suspect DI-only files. The mix should produce more outcome possibilities to our panel than strictly DI-only multiple case files.

4. Confirmation Status

Minimum Requirement: Polygraph results have been confirmed for at least one DI suspect and one NDI suspect *or* polygraph results have been confirmed for two NDI suspects.

Preferred Requirement: Polygraph results have been confirmed for all suspects/witnesses.

Rationale: Some independent method of polygraph result confirmation (usually a confession) is required for establishing ground truth. Sending unconfirmed cases to the panel would be counter to the purpose of the panel study.

5. Criminal Code Type

Minimum Requirement: A criminal investigative case excluding urinalysis cases.

Preferred Requirement: Same as above with the added condition that the overall selection plan include cases from more than one or two criminal code categories.

The Army clearly had the largest database of multiple suspect cases. Potential multiple suspect cases were selected based on priority level with a low score indicating high priority. All criteria listed for a given priority level must be met before that specific priority level is assigned. The criteria for selecting additional Army cases is presented below.

Priority      Criteria

---

1. Number of suspects/witnesses > 2; at least one DI-Confirmed by confession; at least one NDI-Confirmed; at least three confirmed;
2. Number of suspects/witnesses > 2; at least 3 suspects/witnesses with NDI-Confirmed; no DI-Confirmed by confession;
3. Number of suspects/witnesses = 2 or more; (Two types of Category 3 cases)  
Type 1  
-----  
only one DI-Confirmed by confession; only one NDI-Confirmed;  
Type 2  
-----  
only two NDI-Confirmed; no DI-Confirmed by confession;
4. Number of suspects/witnesses > 2; three or more DI-Confirmed by confession; no NDI-Confirmed;
5. Number of suspects/witnesses > 2; only two DI-Confirmed by confession; one or more DI- or NDI-Unconfirmed;
6. Number of suspects/witnesses = 2; two DI-Confirmed by Confession.

**Rules for "Sanitizing" Files**

The present investigators were not permitted access to the actual case files. A procedure was worked out to obtain computer files of the polygraph cases from each branch. PDRI selected the

case files from the computer database and then requested each branch to send the actual case files.

For confidentiality purposes and to meet requirements of the Freedom of Information Act, we requested each military branch was to hire appropriate personnel (with access to the criminal investigative files) to "sanitize" the file by removing all personality identifying information.

The following rules for "sanitizing" files were given to each branch:

1. Photocopy the entire file - investigative reports, polygraph results, actual confessions, and judicial findings if available. Figures, diagrams, and charts should also be photocopied. Photographs should not be photocopied but a note should be made as to their existence. (For example, "2 photographs of the recovered jewels, 1 photograph of the lacerated face of the child", etc.). In addition to photographs, there might be other information that will not be photocopied (e.g., actual x-rays). It is important that any information that has not been photocopied from the original file be documented. Thus, a list will need to be made of items that were not been photocopied. It is possible (if this information is believed to be critical) that this information will be requested at a later point in time.
2. Check the photocopied sheets to ensure that sheets are complete and in the same numerical order as the originals. If the original pages were not numbered, then number the photocopied pages. Make sure both sides of the original pages (back and front) have been photocopied. Make sure all of the photocopied pages are clear and readable. Poor copies need to be replaced. Make sure the photocopied file has the proper case control number (CCN) in bold letters on the cover page of the file.
3. Return the original file to its original location. All of the following changes will be performed on the photocopied file, not the original.

4. On a separate sheet of paper, identify the names of all participants in the case.

5. Assign a code to each type of participant. The codes are as follows:

S1 = Suspect Number 1

S2 = Suspect Number 2

V1 = Victim Number 1

V2 = Victim Number 2

W1 = Witness Number 1

W2 = Witness Number 2

O1 = Other Person Number 1

O2 = Other Person Number 2

Other persons might include a commanding officer, an employer, a character witness, etc. who may make a statement but who are really not suspects, victims, or witnesses to the crime.

If there are more than two suspects, victims, witnesses, or other persons, continue labeling them S3, S4, etc.; V3, V4, etc.; W3, W4, etc.; or O3, O4, etc. Thus, the name of each participant will have a corresponding code that will be used in "sanitizing" the data. It is very important for us to know who said what about whom.

6. On the photocopied file, place the proper code number above the name of the participant each time the name occurs. When codes have been assigned to all names throughout the file, then blacken out the names.

7. Do not erase the age or sex of any persons involved in the case. Do not erase the case control number (CCN).

8. Erase (by blackening out) all social security numbers, military identification numbers, addresses and telephone numbers.

9. No erasures or deletions should be made other than those specified above.
10. When the "sanitized" file is complete and all deletions have been made, photocopy the file again. The second photocopy will ensure that items that were blackened out do not show through. This second photocopied file should be checked to ensure all pages are readable, in numerical order, and complete.
11. Mail the second photocopied file to PDRI. The first photocopied file should be destroyed. The original file should have been refiled prior to any "sanitizing."

#### **Procedure for Organizing and Indexing Polygraph Files**

All military files received were organized and indexed on PDRI premises. The following instructions were provided to individuals hired to organize the files.

"Although the polygraph files we have received from the various branches of military service have been "sanitized" (i.e., identifying information removed), information in the files must remain strictly confidential. You are not to discuss the contents of the files with anyone except the project co-directors."

"The following procedures are to be used in abstracting the case files you have receive. You must work independently."

#### **Procedure**

- I. Information in each file should be reorganized and placed in chronological order as needed. Typically, each piece of information contains a date and time.
- II. Number the pages in the upper right hand corner. Place a square box around the number

so that it will not be confused with other numbers that might be found on the page. Page 1 will always be assigned to the cover page that is included with the abstract. The cover page identifies each subject of the investigation (i.e., S1, S2, etc.), each witness (i.e., W1, W2, etc.), and other participants.

III. The Case Summary should contain the following information in this order:

A. Case descriptive information, including:

1. Case Control Number - found on the cover page
2. Abstracter's name and the date - place in the upper right corner of the first page
3. Service Branch - Air Force, Army, Marines, or Navy. This refers to the military branch which conducted the investigation, not the branch to which the subject(s) of the investigation belong.
4. Criminal Code Category and Definition - See attached list of criminal category definitions. Include both the criminal code category and the definition.
5. Critical/Relevant questions from the polygraph examination - there are usually only two or three of these. They may be found on the polygraph question page or in one of the reports of investigation.

B. Abstract - this should be very brief. The first sentence should contain the criminal charge. You should also include number of suspects, number of witnesses (but no

information about the evidentiary value of their testimony), information about the number of victims or injuries to victims if that is important to the case. The abstract should be general information and you should be careful not to include any conclusions or inferences or judgments of your own. Do not include the results of the polygraph, information about confessions, or results of the investigation in the abstract.

C. Special Terms Used (make two lists) - If an acronym is used only once and defined when it is used, you don't need to include it in the list of special terms. In general, it is probably better to err on the side of including too many terms than too few terms. If a term is not defined in the file, but you can make an educated guess about its meaning, do so but mark your guesses with a question mark. Don't differentiate between legal and illegal drugs, or behaviors, etc. Just define the terms.

1. Military ranks, acronyms, or technical terms - for example, BM3, NAVGPSCOL, BEQ, berthing area.
2. Non-military terms, colloquialisms, non-technical military terms - for example, drug-related terms like "bladder acid", colloquialisms like "higher than a kite", or non-technical military terms like "head" or "rack."

D. Table of Contents - For each piece of information in the case file, include the following:

1. Date (info collected)
2. Number of pages (to insure completeness)
3. Contents - begin with the code that identifies the person making the statement and also note whether this is the person's first, second, third, etc. statement. For example, you might enter "S1's Statement (first)". Note the status of Reports of

Investigation as "pending", "closed", etc.

4. Page number (the new number you placed in a box).

**A sample Table of Contents might be:**

Date	# of pages	Contents	Page
	1	Cover Page	1
10JAN87	3	S1's Waiver and Statement (first)	2
11JAN87	1	S2's Waiver and Statement (first)	5
12JAN87	*	V1's Statement	*
15JAN87	1	Report of Investigation (pending)	7
23FEB87	1	S1 Waiver	8
23FEB87	1	S1 Polygraph Exam Waiver	9

Make a break in the table of contents at the point where the polygraph was given. This gives us an easy way to tell which part of the case file is pre-polygraph and which part is post-polygraph.

If a piece of information described in a Report of Investigation was available to investigators at an earlier date even though no official statement with that date is present in the file, show that in the table of contents. Enter who the information came from and/or what it is, then place an asterisk in the columns for # of pages and for the page number. After the asterisk in place of the page number, put the page number of the Report of Investigation where the information is described and enclose it in parentheses. For example, in the sample table of contents above, the Report of Investigation from 15JAN describes the victim's statement that was made on 12JAN. There is no official statement from V1 on 12JAN but the investigators knew that information on

the day so we want to show that.

If a person signed an Acknowledgment and Waiver of Rights form, but made no statement, call that a Waiver in the table of contents. If they do make a statement, call it a Waiver and Statement in the table of contents.

You may include a listing of the exhibits from the file if there is only one numbering system used in the file. You should be sure to note any exhibits which are *not* included with the file (like photographs) so the panelists won't waste time looking for them.

As you go through the file, whenever an exhibit is referenced, write in the page number in the file where that exhibit is located. For example, if a Report of Investigation refers to Exhibit 1 (the witness' statement), and Exhibit 1 begins on page 3, write "P 3" next to the reference in the Report of Investigation. It is important to write in the "P" because we are going to tell panelists that numbers preceded by a "P" in the file were added by the abstracters. This is the only way panelists will have of knowing that we added numbers to the file.

- E. Missing pages or documents - this should be noted below the Table of Contents. Likewise, if dates were not included on pieces of information, place your estimation of date and reasoning for assigning the date.
- F. Summary of Factual Information - For each piece of information in the file, provide a brief summary of factual information contained. Try to keep it to 1-2 sentences. For example, in S1's statement of 10JAN87, he denied robbing the bank. V1 (12JAN87) identified S1 as the man who gave her the note to hand over the money. Separate the information known about the case prior to the administration of the polygraph exam from the information uncovered subsequent to the exam.

- G. Personal judgments about the case - For each of the subjects in the case, make an overall judgment about whether you think they are guilty or innocent of committing the crime of which they are accused. Then make the same guilty-innocent judgment in relation to each of the relevant questions asked during the polygraph exam. Also judge the confidence you have in your decisions (from .00 to 1.00) and briefly explain your reason(s) for making that decision.
- H. Confessions - Note whether or not any of the subjects confessed specifically to any (or all) of the relevant questions from the polygraph exam. It doesn't count if a subject confessed to a different, a related, or a lesser crime.
- I. Results of the investigation - Tell what action was taken, if the case was dropped, who got charged, etc. This information is usually in the final Report of Investigation.

When typing the above information, you should always start a new page when you start the following sections:

1. Case descriptive information
  2. Table of contents
  3. Summary of factual information
  4. Personal judgments about the case
- J. At the completion of the case, estimate as closely as you can the amount of time you spent on the case.

### **Case Handling Procedures**

These procedures outline how PDRI kept track of which abstracter had which cases and tracked each case as it went from raw, unsorted form to rated and sorted to abstracted (finished).

I. Copies of newly arrived case files

- A. As new cases arrived they were photo-copied and placed in a central safe repository. The cases were color-coded so we could tell at a glance which military branch each case came from.
- B. After photo-copying, each case was placed in a manila folder.

II. Sorting and Rating Cases

- A. The Sorter signed out cases on a master log
- B. The Sorter quickly skimmed the case to ensure completeness of file contents
- C. A sheet of colored paper was placed in the case to divide the information into pre-polygraph, Report of Investigation (ROI), and post-polygraph sections. Within the ROI(s), the information was marked to differentiate pre-polygraph from post-polygraph information. Any confessions that occurred immediately before the polygraph were placed in the post-polygraph section.
- D. The sorter read through the pre-polygraph information (including the pre-polygraph information in the ROI(s)) and made an overall case quality judgment. He or she marked the rating on the case quality form and place the form in the sorted and rated box.

### **III. Organizing and Indexing the cases**

- A. Each abstractor worked with cases that were sorted and rated by someone else.
- B. The abstractor organized and indexed each case using careful instructions.  
He/She filled out an Abstract Summary Form (shown in Figure 2).
- C. The abstractor signed the master log
- D. The abstractor placed the finished case folder back in the central repository.
- E. When an abstractor encountered a word, phase, or an abbreviation unique to a military environment, the definition of this term was placed in a master "glossary of terms".

#### **Case Log**

As cases were received, the tracking of cases was monitored through a case log. The final case log for the project is provided in Appendix M.

#### **Missing Case Files**

Whenever a requested file was declared "missing," PDRI requested an explanation from the service branch. PDRI then requested a suitable replacement file.

The reasons for missing files given by each branch is provided below.

Figure 2

ABSTRACT SUMMARY FORM

Case Control Number

Computer Filename(s)

Service Branch      Date Requested      Date Received      Date Abstracted

Person Abstracting: \_\_\_\_\_ Abstract Procedure: \_\_\_\_\_

Criminal Code: \_\_\_\_\_ Description: \_\_\_\_\_

CASE FILE:

# of Pages: \_\_\_\_\_ # of Subjects: \_\_\_\_\_ # of Others: \_\_\_\_\_

# of Witnesses: \_\_\_\_\_ Abstract Time (min): \_\_\_\_\_

GROUND TRUTH CRITERIA:

Physical Evidence : \_\_\_\_\_

Urinalysis Results: \_\_\_\_\_

A Suspect makes contradictory statements: \_\_\_\_\_

Figure 2 (continued)

## ABSTRACT SUMMARY FORM

## GROUND TRUTH CRITERIA:

Statements of two suspects are contradictory: \_\_\_\_\_

Statements of (Witness, Victim, Other)  
is contradictory to Subject's statement: \_\_\_\_\_

## **Statements of Witnesses, Victims, Others are contradictory:**

Confession to committing crime: S1: \_\_\_\_\_ S2: \_\_\_\_\_ S3: \_\_\_\_\_

OVERALL QUALITY OF EVIDENCE (PRE-CONFESION) (1-10): \_\_\_\_\_

Air Force

1. Files still with our local office, probably still pending in court or waiting to be called for appeal.
2. Have not been received by our records repository from our field units.

Army

1. Polygraph file is available but not investigative file. If the last digit of the case control number is an alphabetic character instead of a number, the investigative file is probably with the Military Police Reports or the Child Offense Service.
2. Files are "unavailable to us".
3. Polygraph file is available but not investigative file. Files are still pending with local office or with local non-military police.
4. "1011 pages not included as per phone conversation with Wolfgang Vinsky".

Navy/Marines

1. The problem seems to be due to the Navy's "automatic tracking system." The tracking system links the CCN (Case Control Numbers) assigned to polygraph cases with a different set of control numbers assigned to the more general criminal cases. Apparently the link between the CCN and the criminal case number is missing hence the problem in locating files (which are filed by the more general code numbers). Files presumably exist and they will send them if and when they find them.

2. "...reasons unknown to us"
3. The above files will not be used as part of the research project because they are old cases.
4. The above files will not be used as part of the research study because they are classified files.
5. The above files will not be used as part of the research study because they can't be located in case control system.
6. The above files will not be used as part of the research study because the CCN numbers are incomplete.

The two principal investigators rated the degree to which a subject's confession statement verified his/her item responses to the polygraph exam. We used the "Quality of Confession Rating Scale" (Figure 3A). Our ratings of 10 polygraph cases included ratings on 56 specific items. The specific items were the relevant questions administered on the polygraph.

Under the assumptions of Shrout and Fleiss's Model 2 (i.e., random sample of k judges, each judge rates each case), the two judge reliability estimate was 0.93.

Both principal investigators reported some difficulty in distinguishing categories 1 and 2 and categories 4 and 5, so these categories were collapsed and a new 0 to 3 rating scale was created. With the new 0 to 3 rating scale, the reliability estimate was 0.98.

Regardless of which scale was used, a high level of inter-rater agreement existed between the principal investigators. Due to the high level of agreement, the subsequent files were reviewed by only the principal investigators. Subsequent "Quality of Confession" ratings were made

Figure 3-A

Quality of Confession Rating Scale

- 5     *Virtually no doubt* that confession verified (or confirmed) this response.
- 4     The confession verified this response *beyond any reasonable doubt*.
- 3     The confession *probably verified* this response.
- 2     The confession *did not verify* this response.
- 1     The confession *clearly did not verify* this response.
- 0     The confession was *unrelated* to the response, *no way of confirming or disconfirming*.
- 0A    The subject confessed to something other than what had been covered by the polygraph questions.
- 0B    The subject confesses to something that may be covered by the polygraph questions but can not be verified because of the terminology being used (e.g., Is the "warehouse" mentioned in the confession the same as "Area 27" mentioned in the polygraph question?)

Categories "0A" and "0B" are sub-categories of Category "0".

independently. The Final "Quality of Confession Rating Scale" is presented in Figure 3B.

#### **Data Analyses - Archival Data**

Due to the nature of the data collected by each service branch, the methods by which the data were stored on the computer, and the lack of standardization of information across branches, the archival data from each branch was analyzed separately.

A series of crosstabulation analyses were conducted to assess the relationship between the results obtained from the Polygraph exam and a series of independent variables. These analyses may help determine, for example, whether suspects with different polygraph results are different on the set of independent variables as well. In a similar manner, another series of crosstabulation analyses assessed the relationship between a set of independent variables and whether or not a confession was obtained. The comparison of cases on these variables will provide preliminary insight into how representative the cases associated with varying degrees of criterion evidence are with respect to the entire sample of military polygraph cases. For example, these analyses may help determine whether suspects who confess are different on the independent variables from subjects who do not confess.

Whether or not a confession is obtained could also be moderated by a series of variables such as military rank, gender, race, educational level, etc.

Figure 3-B

Final "Quality of Confession Rating Scale"

- 3- The confession verified the accusation contained in the relevant question **with almost certainty.**
- 2- The confession **probably verified** the accusation contained in the relevant question.
- 1- The confession **did not verify** the accusation contained in the relevant question.
- NA- Not Applicable (e.g., gender inappropriate question such as asking a male "Did S1 touch your vagina?").
- 0- The confession was **unrelated to the accusation, no way of confirming or disconfirming.**
- OA- The subject confessed to something other than what had been covered by the accusations in the polygraph questions.
- OB- The subject confesses to something that may be covered by the accusations in the polygraph questions but can not be verified because of the terminology being used (e.g., Is the "warehouse" mentioned in the confession the same as "Area 27" mentioned in the polygraph question?)

Categories "OA" and OB" are sub-categories of Category "O".

### **Air Force**

The archival Air Force data collected covered the period from 1986 to 1988. The data files were combined by PDRI to yield a sample of 3981 individuals.

Table 8 presents the results of crosstabulations of independent variables with the dependent variable of Polygraph result (i.e., DI, NDI, Inconclusive/No Opinion). For each independent variable, the value of the chi-square statistic with its corresponding number of degrees of freedom and significance level is provided.

The specific results for each independent variable may be found in Appendix N. For a more thorough examination of the crosstabulation table, the appropriate appendix should be consulted. Frequency distributions for each variable can be obtained by examining the marginal distributions from the crosstabulation tables.

The chi-square values associated with each of the independent variables was significant at the  $p < .001$  level. It is not surprising the chi-square values are statistically significant due to the very large sample sizes.

Some additional information obtained by referring to the corresponding appendices reveals that:

- 1) 68.7% of these Air Force personnel who participated in a polygraph exam were enlisted personnel; 10.2% were civilians.
- 2) 11.5% of the sample requested an exculpatory exam. Of those requesting an exculpatory exam, 80.6% were found to be DI. For those not requesting an exculpatory exam, only 42.0% were found to be DI.
- 3) 12.0% of the sample provided pre-test confessions.

Table 8

## Air Force Data: Crosstabulation of Relevant Variables with Polygraph Results

Variable	Chi-Square			Sample	
	Value	DF	Sig Level	Size	Appendix
Rank	69.08	8	.00000	1669	N-2
Exculpatory Exam	183.68	2	.00000	2827	N-3
Pretest Confession	350.92	2	.00000	28N7	N-4
Post Test Confession	665.25	2	.00000	2827	N-5
Confirmed by Other					
Evidence	200.97	2	.00000	2827	N-6
Type of Crime	236.13	22	.00000	1585	N-7, N-8
Status (i.e. Subject, Victim, Witness)	18.61	4	.00094	2399	N-9
Additional Leads	188.08	2	.00000	2827	N-10
Confession	1023.77	2	.00000	2827	N-11
Confirmed	1133.86	2	.00000	2746	N-12

- 4) 18.1% of the sample provided post-test confessions. 97.9% of those providing post-test confessions had obtained DI results. Only 2.0% of the NDI group made a post-test confession. It may be that investigators were not as rigorous in soliciting post-test confessions from those receiving NDI polygraph results.
- 5) 16.6% of the sample had other (i.e., non-confession) measures of ground truth.
- 6) 31.6% of the sample were accused of crimes against property; 22.9% were accused of crimes against persons.
- 7) 90.1% of the sample were suspects; 6.5% were victims; and 3.4% were witnesses.
- 8) Sometimes the polygraph exam process provides additional leads to resolving the criminal cases. Additional leads were obtained from 10.2% of those tested. Most of the additional leads (i.e., 84.7%) came from DI cases. Regardless of the accuracy of the polygraph, it must be noted that subsequent investigation and interrogation of DI subjects does provide additional leads to resolving the case.
- 9) Confessions (pre, post, or both) were obtained in 26.5% of the cases. 96.4% of the confession-verified cases had DI polygraph results.
- 10) 35.1% of the polygraph cases were confirmed. 90.5% of the confirmed cases had DI polygraph results.

Table 9 presents the results of a similar set of independent variables which were cross-tabulated with the dependent variable of confession obtained (Yes or No).

The chi-square values associated with each of the independent variables, except rank and status, was significant at the  $p < .001$  level. Again, it is not surprising that the chi-squares are statistically significant because the sample size is so large.

Some additional information was obtained by examining Table 9 and the corresponding pages in Appendix N reveals that:

- 1) The rank (e.g. enlisted, officer) of the candidate tested was not related to whether or not a confession was obtained.
- 2) 26.4% of the confessions were confirmed by other evidence. The specific type of "other" evidence is not found on the computer files.
- 3) 78.0% of the additional leads came from individuals who confessed.

### **Army**

The archival Army polygraph data collected covered the period from 1983 to 1989. The data files were combined to yield a sample of 19,804 individuals.

Table 10 presents the results of a series analyses in which independent variables were crosstabulated with the dependent variable "Polygraph Results" (i.e., DI, NDI, Inclusive/No Opinion).

Table 10 shows the general results for each independent variable.

The specific results for each independent variable may be found in a separate pages of Appendix O referred to in Table 10. For a more thorough examination of the crosstabulation table the appropriate pages in Appendix O should be consulted. Frequency distributions for each variable can be obtained by examining the marginal distributions from the crosstabulation tables.

Table 9

## Air Force Data: Crosstabulation of Relevant Variables with Confession

Variable	Chi-Square			Sample	
	Value	DF	Sig Level	Size	Appendix
Rank	8.30	4	.08110	2325	N-13
Exculpatory Exam	18.08	1	.00002	3980	N-14
Confirmation of Results	1168.03	1	.00000	2747	N-15
Confirmed by Other Evidence	189.52	1	.00000	3980	N-16
Type of Crime	152.17	11	.00000	1587	N-17, N-18
Status (i.e. Subject, Victim, Witness)	3.36	2	.18601	3044	N-19
Additional Leads	710.25	1	.00000	3980	N-20

Table 10

## Army Data: Crosstabulation of Relevant Variables with Polygraph Results

Variable	Chi-Square			Sample	
	Value	DF	Sig Level	Size	Appendix
Confirmation	3261.57	4	.00000	17527	O-2
Confession	7350.58	2	.00000	19804	O-3
Rank	446.91	16	.00000	19804	O-4
Polygraph Exam					
Technique	215.03	18	.00000	17812	O-5, O-6
Pre-test Confession	1757.28	2	.00000	19804	O-7
Post-Test Confession	4420.31	2	.00000	19804	O-8

Due to the very large sample sizes, the chi-square value associated with each of the independent variables was significant at the p<.00001 level.

Some additional information obtained by referring to the corresponding pages in Appendix O reveals that:

- 1) The polygraph results of 70.8% of those tested were listed as "confirmed". Of the confirmed cases, 87.6% were DI-confirmed and 12.4% were NDI-confirmed.
- 2) The polygraph results of 43.0% of those tested were confirmed by a confession. Confessions could be obtained pre-polygraph, post-polygraph, or both. 64.2% of the DI cases yielded confessions. No confessions were obtained from NDI cases.
- 3) 83.1% of polygraphed individuals were enlisted personnel; 2.1% were officers; 3.1% were non-government civilians; and 3.0% were military family members.
- 4) The Zone of Comparison (ZOC) technique was used in 42.4% of the Army polygraph cases; 39.7% of the cases used the MGQT technique; and 17.4% of the cases used a combination ZOC-MGQT technique.
- 5) A pre-test confession was obtained from 15.3% of the cases.
- 6) Post-test confessions were obtained from 31.2% of the cases. Only individuals with DI results made post-test confessions.

Table 11 presents the results of two independent variables: "Rank" and "Polygraph Exam Technique" which were been crosstabulated with the dependent variable "confession obtained". The chi-square value associated with the "Rank" variable was significant at p<.00001 level. The chi-

Table 11

## Army Data: Crosstabulation of Relevant Variables with Confession

Variable	Chi-Square			Sample	
	Value	DF	Sig Level	Size	Appendix
Rank	118.84	8	.00000	19804	O-9
Polygraph Exam Technique	26.57	9	.00165	17812	O-9, O-10, O-11

square value associated with the "Polygraph Exam Technique" variable was not significant at  $p<.001$  level.

Some additional information obtained from examining Table 11 and the corresponding pages in Appendices O reveals that:

- 1) 44.4% of the enlisted personnel who were polygraphed had made a confession; 35.2% of the officers who had been polygraphed had confessed; and 37.3% of the polygraphed non-government civilians had confessed.
- 2) The polygraph exam technique used is not related to whether or not a confession was obtained.

Table 12 presents a series of one-way analyses of variance. The dependent variable under consideration was "Length of Polygraph Exam (in minutes)". One-way analysis of variance yielded significant results ( $p<.0001$ ) for each of the independent variables, namely, the Polygraph Results, whether or not a confession was obtained, and whether or not a pre-test (or post-test) confession was made.

Additional information from the corresponding pages in Appendices O referred to in Table 12 revealed:

- 1) For individuals with a DI outcome, the mean length of the polygraph exam was 203 minutes; for individuals with an NDI outcome the mean length was 166 minutes; for individuals with an Incomplete/No Opinions outcome the mean length was 181 minutes.
- 2) The length of the polygraph exam was longer for individuals who confessed (mean = 211 minutes) than for those who did not confess (mean = 179 minutes).

Table 12

Army Data: One-way Analysis of Variance: Length of Polygraph Exam  
(in Minutes) by Relevant Variables

Variable	F- Value	Sig Level	Size	Sample Appendix
Polygraph Result	245.03	.0000	19800	O-12, O-13
Confession	485.73	.0000	19800	O-14
Pre-Test Confession	523.54	.0000	19800	O-15
Post-Test Confession	2249.25	.0000	19800	O-16

- 3) The polygraph exam took less time for individuals who made a pre-test confession (mean = 155 minutes) than it did for those who did not (mean = 200 minutes).
- 4) The polygraph exam took longer for individuals who made post-test confessions (mean = 241 minutes) than for those who did not (mean = 171 minutes).

### Marines

The archival Marine data collected covered the period 1987 to 1989. The original data were handwritten in a logbook. PDRI had the data from the logbook entered into a computer file for analyses. In the Marine polygraph data set 842 individuals participated in a polygraph exam. Table 13 presents the results of a series of analyses in which independent variables were cross-tabulated with the dependent variable "Polygraph Result". For each independent variable, the value of the chi-square statistic with its corresponding number of degrees of freedom and significance level is provided in Table 13.

The specific results for each independent variable may be found in a separate pages of Appendix P. For a more thorough examination of the crosstabulation table, the appropriate appendix should be consulted. Frequency distributions for each variable can be obtained by examining the marginal distributions from the crosstabulation tables. The chi-square value associated with each of the independent variables was significant at the  $p < .001$  level.

Some additional information obtained by referring to Table 13 and the corresponding appendices reveals that:

- 1) 71.7% of the polygraph cases were DI; 20.0% were NDI; and 8.2% were Inconclusive/No Opinion.

Table 13

## Marine Data: Crosstabulation of Relevant Variables with Polygraph Results

Variable	Chi-Square			Sample	
	Value	DF	Sig Level	Size	Appendix
Type of Crime	74.16	10	.00000	644	P-2
Exculpatory Exam	36.25	2	.00000	623	P-3
<b>Polygraph Exam</b>					
Technique	48.18	20	.00040	545	P-4, P-5
<b>Pre/Post Test</b>					
Confession	100.15	6	.00000	631	P-6, P-7

- 2) 59.0% of the polygraph cases involved crimes classified as "Special Activities"; 24.7% were classified as "Crimes Against Persons"; 5.4% were classified as "Criminal Sexual Behavior" and 4.3% were classified as "Crimes Against Property".
- 3) DI-results were obtained from 90.6% of those classified as "Crimes Against Persons"; 77.1% of "Criminal Sexual Behavior" crimes; 67.1% of the crimes classified as "Special Activities", and 64.3% of the "Crimes Against Property".
- 4) 35.3% of the polygraph cases were exculpatory exams.
- 5) The polygraph technique employed was the ZCT for 67.0% of the cases with 16.1% of the cases using the MGQT technique.
- 6) No confession was obtained from 66.1% of the individuals. Pre-test confessions were obtained from 5.1% of the cases and post-test confession was obtained from 24.4% of the individuals.
- 7) 99.4% of cases where a post-test confession was obtained had a DI-result, 60.2% of the "no confession" cases had a DI result.

Table 14 presents the results of a similar set of independent variables which crosstabulated with the dependent variable "Confession/Admission Obtained (Yes/No)".

The chi-square value associated with each of the independent variables was significant at the p<.0001 level.

Some additional information obtained by examining Table 14 and the corresponding pages form Appendix P reveals that:

Table 14

Marine Data: Crosstabulation of Relevant Variables with Confession Results

Variable	Chi-Square			Sample	
	Value	DF	Sig Level	Size	Appendix
Type of Crime	39.68	5	.00000	837	P-8
Exculpatory Exam	23.31	1	.00000	739	P-9
Polygraph Exam					
Technique	36.52	10	.00007	631	P-10, P-11
Polygraph Results	101.49	2	.00000	645	P-11

- 1) 41.8% of the "Crimes Against Persons" cases were confirmed by confession or Admission; whereas only 26.5% of the "Criminal Sexual Behavior" cases and 13.5% of the "Crimes Against Property" cases were confirmed by confession or admission.
- 2) 39.1% of the exams conducted for exculpatory purposes were confirmed whereas only 22.4% of the nonexculpatory exams were confirmed.
- 3) 38.2% of the cases using the MGQT technique were confirmed; 28.3% of the cases employing the ZCT technique were confirmed.
- 4) 97.2% of the confirmed cases were DI-Confirmed.

#### Navy

The archival Navy data collected covered the period from 1980 to 1987. The data files were combined to yield a sample of 9,834 individuals.

Table 15 presents the results on a series of analyses in which independent variables were cross-tabulated with the dependent variable "Polygraph Results". For each independent variable the value of the chi-square statistic with its corresponding number of degrees of freedom and significance level is provided in Table 15.

The specific results for each independent variable may be found in separate pages of Appendix Q. For a more thorough examination of the crosstabulation Table, the appropriate Appendix should be consulted. Frequency distributions for each variable can be obtained by examining the marginal distributions from the crosstabulation tables.

Table 15

Navy Data: Crosstabulation of Relevant Variables with Polygraph Results

Variable	Chi-Square			Sample	
	Value	DF	Sig Level	Size	Appendix
Gender	30.31	2	.00000	9567	Q-2
Race	167.95	4	.00000	9129	Q-3
Education Level	119.23	6	.00000	9042	Q-4
Confession	2924.70	4	.00000	9659	Q-5
Exculpatory Exam	1525.00	2	.00000	9433	Q-6
Lawyer	94.54	2	.00000	5635	Q-7
Type of Crime	2849.69	148	.00000	9396	Q-8 to Q-14
Confirmation	3355.23	4	.00000	9396	Q-15
Status (i.e., Subject, Victim, Witness, Co-subject)	197.66	6	.00000	1386	Q-16
Rank	585.44	18	.00000	9427	Q-17, Q-18

Table 16

Navy Data: Crosstabulation of Relevant Variables with Confession

Variable	Chi-Square			Sample Size	Appendix
	Value	DF	Sig Level		
Gender	19.87	2	.00005	9736	Q-19
Race	17.90	4	.00129	9181	Q-20
Education Level	39.62	6	.00000	9050	Q-21
Exculpatory Exam	570.21	2	.00000	9600	Q-22
Lawyer	200.33	2	.00000	5650	Q-23
Type of Crime	1124.25	148	.00000	9830	Q-24 to Q-30
Confirmation	7019.63	4	.00000	9403	Q-31
Status (i.e., Subject, Victim, Witness, Co-subject)	42.57	6	.00000	1386	Q-32
Rank	174.81	18	.00000	9583	Q-33, Q-34

The chi-square values associated with each of the independent variables was significant at the  $p < .00001$  level. Note, however, the very large sample size. Some additional information obtained by referring to the corresponding pages in Appendix Q listed in Table 15 reveals that:

- 1) 91.4% of the individuals polygraphed were males
- 2) 68.6% of the individuals tested were "White"; 29.2% were "Black"; and 2.2% were "Other".
- 3) 65.2% of the individuals tested were high school graduates; 12.5% were high school dropouts. 17.2% had some post high school training; and 5.1% were college graduates.
- 4) Post-test confessions were obtained from 25.0% of the individuals tested; 10.7% of the individuals tested made pre-test confessions; and 64.3% of the polygraphed individuals made No Confession.
- 5) 99.0% of the post-test confessions were obtained from DI cases whereas only 41.9% of the "No Confession" cases had DI results.
- 6) 59.7% of the polygraph cases were conducted for exculpatory purposes.
- 7) A lawyer's assistance was requested for 19.0% of the cases.
- 8) 40.3% of the cases were Confirmed; 59.3% were Unconfirmed; and 0.3% were Contradicted.
- 9) Of the Confirmed cases, 98.1% were DI-Confirmed and 1.7% were NDI-Confirmed.
- 10) Of the Unconfirmed cases, 39.3% had DI results and 53.9% had NDI results.
- 11) Of the individuals tested, 83.2% were classified as "Subject"; 0.4% as "Victims". 6.1% as "Witness"; and 10.3% as "Co-Subject".
- 12) 87.0% of the data collected from the Navy were from "Enlisted" personnel; 3.1% from "officers"; and the remainder were other civilians.

Table 16 presents the results of a similar set of independent variables which were crosstabulated with the dependent variable "Confession" (i.e., O = Post-test confession; P = Pre-test confession; and X = No confession).

The chi-square value associated with each of the independent variables (except Race) was significant at the p<.0001 level.

Some additional information obtained by examining Table 16 and the corresponding pages in Appendix Q reveals that:

- 1) The race of the candidate tested is not related to whether or not a confession was obtained.
- 2) No confessions were obtained from 74.8% of the college graduates, 64.7% of those with some post-high school training, 62.3% of the high school graduates, and 62.8% of the high school dropouts.
- 3) Only 17.7% of those who requested the assistance of a lawyer made a confession, whereas 40.6% of those who did not request a lawyer made a confession.
- 4) 86.8% of the confirmed cases were from individuals making pre or post-test confessions. The remainder of the cases were confirmed by other means.

Table 17 contains the results from two one-way analyses of variance. The dependent variable was "Length of the Polygraph Exam" (in minutes). The independent variable for the first one-way ANOVA was "Polygraph Result". The independent variable for the second one-way ANOVA was "Confession".

Table 17

Navy Data: One-way Analysis of Variance: Length of Polygraph Exam  
(in Minutes) by Relevant Variables

Variable	F- Value	Sig Level	Size	Sample Appendix
Polygraph Result	2.30	.1005	9659	Q-35, Q-36
Confession	131.76	.0000	9830	Q-37, Q-38

Table 17 and the corresponding pages in Appendix Q listed in the table show:

- 1) No significant mean difference found between length of the polygraph exam and the result of the polygraph exam. The average length of the polygraph exam was 113.5 minutes.
- 2) A significant difference ( $p < .0001$ ) in length of the polygraph exam between groups making confessions and those not making confessions. The polygraph exam took longer for individuals who confessed (mean time with pre-test confession = 146 minute; mean time with post-test confession - 113 minutes) than for those who did not confess (mean time = 106 minutes).

Table 18 contains the results from two one-way analysis of variance using the dependent variable "Length of the Post-test Exam". The independent variable for the first one-way ANOVA was "Polygraph Result". The independent variable for the second one-way ANOVA was "Confession".

Table 18 and the corresponding pages in Appendix Q listed in the table reveals that:

- 1) A significant difference in post-test exam length for the different Polygraph Result groups. The mean length of the post-test exam was 74.4 minutes for the DI group; 25.5 minutes for the Inconclusive/No Opinion Group; and 12.4 minutes for the NDI group.
- 2) Significant differences in post-test exam length existed for the different "Confession" groups. The mean length in the post-test exam was 13.0 minutes when a pre-test confession was made; 31.1 minutes if no confession was made; and 118.2 minutes if a post-test confession was obtained.

Table 18

Navy Data: One-way Analysis of Variance: Length of Post-test Exam by Relevant Variables

Variable	F- Value	Sig Level	Sample Size	Appendix
Polygraph Result	327.17	.0000	9659	Q-39, Q-40
Confession	630.20	.0000	9830	Q-41, Q-42

Table 19 contains the results from two one-way analysis of variance using the dependent variable "Length of Chart Time". The independent variable for the first one-way ANOVA was "Polygraph Result". The independent variable for the second one-way ANOVA was "Confession".

Table 19 and the corresponding pages in Appendix Q listed in the table reveals that:

- 1) A significant difference ( $p<.001$ ) in "Length of Chart Time" for the different Polygraph Result groups. The mean length of chart time was 14.0 minutes for the Inconclusive/No Opinion group; 15.9 minutes for the DI group; and 18.6 minutes for the NDI group.
- 2) A significant difference ( $p<.0001$ ) in "Length of Chart Time" between groups making confessions and those not making confessions. The mean chart time was 4.2 minutes for individuals making pre-test confessions; 16.8 minutes for individuals not making confessions; and 20.4 minutes for those making post-test confessions.

Table 19

Navy Data: One-way Analysis of Variance: Length of Chart Time by Relevant Variables

Variable	F- Value	Sig Level	Sample Size	Appendix
Polygraph Result	7.34	.0007	9659	Q-43, Q-44
Confession	75.24	.0000	9830	Q-45, Q-46

## PANEL STUDY

A major purpose of conducting a panel study is to establish an alternate measure of ground truth. The present study examined the feasibility of using the decisions of panels to establish ground truth.

The present study has attempted to minimize some of the methodological problems found in previous studies, namely:

1. Panel studies to date are problematic in that it is unclear how much and what types and quality of data the panels had to work with; and
2. Previous panel studies have access to confessions elicited from only failed polygraph tests (i.e., DI-confirmed cases). These confessions are associated with an unrepresentative set of charts and are not independent of the outcome of the polygraph test.

Regarding the first problem, it is imperative that the case data presented to the panel be as complete and accurate as possible. Often the case data is abstracted prior to presentation to the panel. The process of abstracting files may contaminate the panel review process since the pertinent facts presented in the abstract have been "filtered" by some person's idiosyncratic views of what is and what is not important. Minimally, the material presented to the panel, if it is abstracted, must be gathered with demonstrated reliability so that the relevant facts of the case are not "filtered". The process of abstraction must be clearly elaborated with explicitly defined criteria as to what constitutes the relevant facts of the case. In addition, of special concern is whether or not panels have access to confessions elicited following a failed polygraph test. To the extent they do have these confessions, the panel will most likely affirm the confession, and the distinctiveness of these two methods for establishing ground truth becomes blurred.

To address the above mentioned problems, the present study:

1. controlled the amount of information provided to the panel. For example, the "partial" level presents panel members with all case data placed in chronological order, up to the point in time the polygraph test was administered (excluding of course the polygraph charts and confessions). Thus, ratings obtained from panel members were not contaminated by knowledge of confession statements and
2. provided all case information in its original form to the panel members. This avoided the "filtering" of information by an abstractor.

Regarding the second problem mentioned above, confession studies have been criticized because confessors are a select group. While this is true, a more serious problem with confession studies is that the polygraph charts associated with confessions represent a select group (Iacono, 1989; Iacono & Patrick, 1987, 1988). In the usual case, confessions are obtained when the examiner attempts to elicit them from a subject whose charts are indicative of deception. Thus, confessions will be obtained only from those who produced deceptive charts. Individuals who generate nondeceptive charts will not be asked to confess. While many of these people will be innocent, some will be guilty persons who erroneously produced truthful-looking charts. The charts associated with these "false negative" errors will never become part of a validity study when confessions are used to establish ground truth. Likewise, an unknown number of innocent people will erroneously produce deceptive charts. However, these individuals will generally not confess to an act they did not commit, and these "false positive" errors will never become a part of the validity study. Note that if some innocent people do confess following a failed polygraph test, this type of error is incorrectly counted as a hit in validity studies using confessions.

Thus, the major problem with confession studies is that almost all of the cases wherein the original examiner made an error are systematically excluded from the data sample. The only cases included are those where both the charts indicate deception and the examinee confesses. Since

numerical scoring of polygraph charts is highly reliable, giving these charts of confession verified guilty and innocent subjects to other examiners to rescore blindly will of course confirm the original examiner's assessment. As a consequence, in confession studies, because cases where potential errors could have been made are eliminated, inaccurate estimates of polygraph validity will be obtained. Accuracy rates for guilty individuals will be especially overly optimistic and should be close to 100% when numerical scoring is used even if the polygraph test had zero validity. Interpreting the data from innocent subjects will also be misleading for these same reasons.

The arguments developed above also apply to panel studies in which the panels have access to confessions elicited following failed polygraph tests. A confession is likely to be persuasive evidence to a panel. However, because confessions will be associated with an unrepresentative set of charts, confessions elicited following failed examinations would not be presented to the panel if a legitimate estimate of polygraph validity is desired.

As with other reported studies, the preponderance of confession-verified cases in the present study collected from the military archives are DI-confirmed. Thus, for the most part, confessions are only obtained when deceptive charts are produced. In the present study, several design improvements were incorporated to minimize the problems of unrepresentativeness in the cases reviewed by the panels:

1. Only multiple-suspect case files were reviewed by our panels. The multiple-suspect files selected contained a greater mixture of DI- and NDI-confirmed cases. Thus, with multiple suspect cases, panel members are not limited to reviewing only DI-confirmed cases or confessions elicited from failed exams. For example, a rape case would have two individuals who have taken the polygraph test where only one of the them told the truth. Ratings were obtained from the panel members for every suspect/victim in the multiple-suspect case file.

2. The panel members in the present study were not aware of which of the multiple suspects rated also had confession data, which subjects rated had been administered the polygraph exam, nor were they aware of the number of suspects who were guilty or not guilty. With so many unknowns, panels members could not conclude that most suspects were DI and make ratings accordingly.

### Description of Panel Members Selected

The panels must be sufficiently large so as to be generally representative of the type of expertise they are trying to tap. One efficient way of doing this (rather than setting up dozens of panels) is to have, for example, 18 members on a panel and divide them into subpanels of varying sizes. For example, one could examine the data collected from three-person panels by analyzing the outputs of all panels with combinations of three members. The result would be an overall estimate of the ability of the panel plus estimates of the range of results achieved by subpanels of three individuals each. One could also vary the size of the subpanels and determine the amount of agreement among subpanel members and the accuracy of the decisions.

Three types of panels are required by the RFP: those composed of members with legal training, police training, and no special expertise (i.e., lay persons).

Thus, three panels were selected, each consisting of six members. One panel included lay persons, a second included police officers, and a third included attorneys.

Prior to being selected to serve on a panel, prospective panelists were given a brief overview of the nature of the rating task and the types of specific incident criminal case files to be reviewed. One prospective (lay) panelist chose not to participate. This participant reported that she felt uncomfortable about the possible graphic nature of file content. Another lay panelist was selected in her place.

All lay persons selected were registered voters residing in the Minneapolis/St. Paul Metropolitan Area. All lay panel members could conceivably be selected to appear as jurors in civil courts.

The police officer panel was selected from officers with at least five years of criminal investigative experience. The St. Paul Police Department was requested to provide officers from their investigative unit. Police panel members were then selected from that unit. Unfortunately there were no female police officers who met our selection criteria.

The Minnesota Trial Lawyers Association provided a list of attorneys that might be selected for the panel. Lawyers residing in the Twin City Metropolitan area were selected from this list. Only trial lawyers with a minimum of five years trial experience were selected.

Table 20 provides details of each panel's composition relative to gender, age, race, education, and experience. Appendix R provides copies of the demographic questionnaires used to collect this data. All panel members were Caucasian. The police officer panel had a mean 22.5 years of experience as police officers and were involved with criminal investigative work for a mean of 10.8 years. Four of the six police officers had previous military experience. The attorney panel, on average, had passed the bar exam nine years ago and had a mean of nine years experience as a trial lawyer. All panel members signed confidentiality statements in which they agreed not to discuss the case contents or disclose any case information to others.

### **Selection of Cases for Panels**

Only multiple-suspect criminal investigative cases were selected for panel review. By selecting multiple suspect cases, a greater opportunity for selecting DI and NDI-confirmed subjects was made possible. This would not be possible if only single-suspect cases were selected since the vast majority of these confirmed cases were DI-confirmed.

Table 20  
Panel Composition: Gender, Age, Race, Education, and Experience

	Panel Type			
	Lay Persons	Police Officers	Lawyers	Total
Number of Members	6	6	6	18
Gender Composition	2 males 4 females	6 males	3 males 3 females	11 males 7 females
<i>Age:</i>				
Mean	27.83	45.17	37.50	36.83
S.D.	3.66	1.60	7.26	8.57
Race	6 Caucasian	6 Caucasian	6 Caucasian	18 Caucasian
Educational Level	1 "vo-tech deg." 1 "some college" 3 "bach. deg." 1 "some grad work"	5 "some college" 1 "MA degree"	6 "law degree"	1 "vo-tech deg." 6 "some college" 3 "bach. deg." 1 "some grad wk" 1 "MA degree" 6 "law degree"

#### Police Officer Qualifications

# Years as Police Officer:

Mean .....	22.50
S.D. ....	1.98
Range .....	20-24

# Years in Criminal Investigation:

Mean .....	10.83
S.D. ....	2.64
Range .....	7-15

#### Lawyer Qualifications

# Years Since Passed Bar:

Mean .....	9.00
S.D. ....	6.90
Range .....	5-23

# Years of Trial Experience:

Mean .....	9.00
S.D. ....	6.90
Range .....	5-23

The selected case files needed to meet some criterion of ground truth. The most common criterion for ground truth is a confession or admission of guilt.

The senior author and consultant (hereinafter referred to jointly as "principal investigators") reviewed the case files prior to presenting them to the panels to ensure cases selected met a ground truth criterion (i.e., confession-verified cases). For each case the investigator was given:

1. the set of relevant polygraph questions administered throughout the course of the investigation, and
2. the confession(s) and/or admission statement(s) obtained throughout the course of the investigation.

Each investigator was then asked to:

1. determine whether each polygraph question was confirmed by the confession, and
2. rate how certain they were that the polygraph question was verified by the confession.

For cases which involved administration of multiple polygraph exams, the polygraph questions reviewed included the unique (i.e., non-overlapping) set of polygraph questions and included confession and/or admission statements from all persons polygraphed (if they existed). Confessions or admissions of guilt from non-polygraphed subjects were included as well.

All files were reviewed and evaluated for "Quality of Confession." (Refer to Figure 3B, page 72.) Only confession-verified cases (where both investigators agreed with a rating of "3") were provided to the panel members. This constituted our Ground Truth criteria. Several larger confession-verified files (100 pages plus) were excluded from the study because these files would have required more time than was available for panel review.

## Description of Rating Forms

Panel members were presented with case material corresponding to two points in time. The level of information provided for each point in time varied. Time 1 materials contained the case investigative information collected *prior* to the first polygraph examination administered. Time 2 materials contained all investigative information (excluding polygraph results) collected until the case was closed. Ratings were made on all of the material presented in each time period.

Four rating forms and two rating sheets were completed independently by each panel member for each case. A sample set of rating forms for one case may be found in Figure 4A and 4B. A brief description of each form in the order they were administered follows:

***Global Rating 1.*** This form was used to provide a global rating for each subject in the investigation. Each panel member reviewed all the pre-polygraph case material prior to completing this form. The panel member then made a judgment as to whether or not, and to what level of certainty, a specified subject committed a crime based on the pre-polygraph case information. Refer to Figure 4A for a sample Global Rating 1 Form.

***Initial Rating Form (Specific Questions).*** This form was used to provide ratings on specific questions for each subject in the investigation. The specific questions on this form corresponded to the relevant questions asked during the polygraph examination(s). Panel members were not aware which relevant questions applied to which suspect. In general, ratings were made for each suspect/victim for each relevant question regardless of whether or not the relevant question was originally asked of that person. In some cases a specific question asked of one suspect would not be relevant to another suspect and the panel was informed not to make a specific rating. For example, a rating would not be made for a gender inappropriate question such as asking a male "Did S1 touch your vagina?" The panel member made a judgment as to whether or not the pre-polygraph case information verified the specific accusation raised in each relevant question.

Refer to Figure 4B for a sample Initial Rating Form.

Figure 4A

Envelope 1 - Global Rating 1  
Case Number 1

Rater

Global Rating Form

Service Branch: Air Force

CCN:

Crime: Child Abuse

A. Did S1 commit the above crime?

- \_\_\_\_\_ *Definitely did not* commit the above crime  
\_\_\_\_\_ *Probably did not* commit the above crime  
\_\_\_\_\_ *Undecided*  
\_\_\_\_\_ *Probably did* commit the above crime  
\_\_\_\_\_ *Definitely did* commit the above crime

B. Did S2 commit the above crime?

- \_\_\_\_\_ *Definitely did not* commit the above crime  
\_\_\_\_\_ *Probably did not* commit the above crime  
\_\_\_\_\_ *Undecided*  
\_\_\_\_\_ *Probably did* commit the above crime  
\_\_\_\_\_ *Definitely did* commit the above crime

C. Did O1 commit the above crime?

- \_\_\_\_\_ *Definitely did not* commit the above crime  
\_\_\_\_\_ *Probably did not* commit the above crime  
\_\_\_\_\_ *Undecided*  
\_\_\_\_\_ *Probably did* commit the above crime  
\_\_\_\_\_ *Definitely did* commit the above crime

Figure 4B

Envelope 1 - Initial Rating  
Case Number 1

Rater

**Rating Form**

Service Branch: Air Force

CCN:

Crime: Child Abuse

Individuals

Questions	S1	S2	O1
	<u>Ratings</u>	<u>Ratings</u>	<u>Ratings</u>
1. Did you ever intentionally shake V1?	_____	_____	_____
2. Did you ever shake that baby?	_____	_____	_____
3. Did you ever violently shake that baby?	_____	_____	_____
4. Did you ever deliberately violently shake that baby?	_____	_____	_____
5. Did you ever violently shake that baby in a fit of anger?	_____	_____	_____
6. Did you ever deliberately violently shake that baby in a fit of anger?	_____	_____	_____
7. Did you ever choke that baby's throat?	_____	_____	_____
8. Did you ever intentionally choke V1's throat?	_____	_____	_____

The rating scale used in rating the specific questions is as follows:

I believe the information contained in the case file.....

**3 verified the accusation with almost certainty**

**2 probably verified the accusation**

**1 did not verify the accusation**

**0 was unrelated to the accusation or there was no way to confirm the accusation**

***Final Rating Form (Specific Questions).*** This form was completed after *all* the case information (except for polygraph results) had been reviewed. Additional information in the case file included additional statements, confessions, and case material collected after the first polygraph exam had been administered. This form is identical to the Initial Rating Form. (Refer to Figure 4B for sample form.) The only difference between the two forms is the level of information provided prior to responding to the specific questions. With the Final Rating Form, the panel member made a judgment as to whether or not all the evidence collected in the case verified the specific accusation raised in each question.

***Global Rating 2.*** This form was also completed after *all* the information (except for polygraph results) had been reviewed. This form is identical to the Global Rating 1 Form. (Refer to Figure 4A for sample form.) The only difference between the two forms is the level of information provided to the panel member prior to responding to the global questions. The panel member made a judgment as to whether or not, and to what level of certainty, a specified subject committed a crime based on all the information presented in the case.

Appendix S contains the Global Rating 1 Form and the Initial Rating Form for each of the 27 cases evaluated in the study. Since the Global Rating 2 and the Final Rating Forms are identical in content to the Global Rating 1 and Initial Rating Forms, these two additional forms were not included in Appendix S.

## Description of Cases Used

Twenty-seven confession-verified multiple suspect case files were selected for the panels. Table 21 provides details of the case files relative to the types of crimes covered, the size of the files, the number of suspects, the number of crimes, and the number of ratings made.

For example, by referring to Table 21 and Figures 4A and 4B, we see that case number 1 was a child abuse case whose case file contained 50 pages of statements, investigative reports, supplementary evidence, and other information. This first case had three suspects: S1, the father; S2, the mother; and O1, the baby sitter; any one or all of whom presumably could have committed the crime of child abuse. Since case 1 involved only one crime, the number of suspects and the number of unique suspects are the same. For case number 21, however, we have two crimes, housebreaking and false statement, with two unique suspects, S1 and S2, each accused of both crimes. Thus, case number 21 had four suspects, but in reality only two (unique) individuals were involved. Referring again back to case number 1, each panel member was required to make three Global I ratings (Figure 4A); three Global II ratings (same as Global I ratings except the data is collected at a later point in time); eight initial ratings per unique suspect (Figure 4A), and eight final ratings per unique suspect (same as initial ratings except the data is collected at a later point in time). Thus, the total number of ratings made by each panel member for case number 1 is  $3 + 3 + (3 \times 8) + (3 \times 8) = 54$  ratings.

In all, 1062 individual ratings were made by panel members on 71 suspects. Each case had from two to four suspects. Some suspects were accused of multiple crimes, thus the number of unique individuals who were suspects was 62. The average size of the case files reviewed was 42 pages.

Table 21

## Panel Case Descriptions: Number of Pages, Suspects, Crimes, and Ratings

Case Number	Crime(s)	Size (# of pages)	# of Suspects	# of Unique Suspects	# of Crimes	Number of Ratings				
						Global Ratings I	Global Ratings II	Specific Ratings Initial	Specific Ratings Final	Total
1	Child Abuse	50	3	3	1	3	3	24	24	54
2	Child Abuse	68	2	2	1	2	2	14	14	32
3	Property Destruction	40	2	2	1	2	2	8	8	20
4	Homosexual Activity	36	2	2	1	2	2	26	26	56
5	Child Abuse	49	2	2	1	2	2	10	10	24
6	Larceny	31	2	2	1	2	2	10	10	24
7	Bribery/Removing a Public Record	37	4	2	2	4	4	18	18	44
8	Housebreaking	49	2	2	1	2	2	14	14	32
9	False Official Statement/Fraud	47	2	2	1	2	2	6	6	16
10	Larceny, Government	38	4	4	1	4	4	44	44	94
11	Housebreaking	34	2	2	1	2	2	12	12	28
12	False Swearing/Indecent Acts	42	3	3	1	3	3	14	14	34
13	Housebreaking	61	3	3	1	3	3	30	30	66
14	Rape/Sodomy/False Swearing	28	3	2	3	3	3	9	9	24
15	Aggravated Assault	59	2	2	1	2	2	14	14	32

Table 21 (continued)

Case Number	Crime(s)	Size (# of pages)	# of Suspects	# of Unique Suspects	# of Crimes	Number of Ratings					
						I	II	Global Ratings	Initial Final	Specific Ratings	Total
16	Forgery/False Swearing	36	3	2	2	3	3	3	14	14	34
17	Larceny, Private/False Swearing	28	4	3	2	4	4	4	12	12	32
18	Larceny	37	2	2	1	2	2	2	18	18	40
19	Housebreaking	52	3	3	1	3	3	3	27	27	60
20	Arson	56	2	2	1	2	2	2	16	16	36
21	Housebreaking/False Statement	40	4	2	2	4	4	4	16	16	40
22	Indecent Assault/False Swearing	29	3	2	2	3	3	3	7	7	20
23	Larceny/False Swearing	23	3	2	2	3	3	3	14	14	34
24	Larceny/False Swearing	50	3	3	1	3	3	3	40	40	86
25	Larceny, Government Property	46	2	2	1	2	2	2	18	18	40
26	Larceny, Government Property	47	2	2	1	2	2	2	14	14	32
27	Sodomy	25	2	2	1	2	2	2	12	12	28
<b>Total</b>		<b>1138</b>	<b>71</b>	<b>62</b>	<b>35</b>	<b>71</b>	<b>71</b>	<b>461</b>	<b>461</b>	<b>1062</b>	

## **Data Collection Procedure**

Panel members reviewed all case files and completed the rating tasks at the office of Personnel Decisions Research Institutes, Inc. (PDRI). This procedure ensured the physical security of the files, prevented any potential attempts to duplicate file contents, and helped ensure confidentiality of case materials was maintained. Upon arrival at PDRI, panel members were provided a verbal overview of the project's objectives and were given the "Case Review Process" handout (Figure 5) and an initial set of five large envelopes containing case materials. The investigator then read the "Case Review Process" procedure aloud, demonstrated the process without actually completing the rating forms, and answered questions. All panel members were informed to work independently and not to communicate with anyone else about the contents of the files or the rating process. Prior to participating, all panel members had previously signed confidentiality agreements as well. While at PDRI, panel members were strictly monitored during the rating process and a PDRI staff member was always present. Although all panelists reviewed the same pre-selected 27 case files, the panelists started with different case numbers and completed the cases in different orders. Five complete sets of case materials were available. Rating forms and completed case folders were collected after one to three cases were completed. New case materials were then given to replace the completed files. The rating forms were checked immediately, after completion to insure all ratings were made. The case review process took 12-16 hours to complete. Panel members completed the ratings over a 2 to 3 day period of time. All panel members were paid for their work.

## **Panel Study: Statistical Analyses and Results**

All data were keypunched, verified for coding accuracy by a different individual, and processed using the SPSS statistical package.

Figure 5  
Case Review Process

**A. Materials**

Each panel member will receive the following materials:

1. Each criminal case file is contained in a separate large envelope. On the outside of each large envelope is printed the case number, the service branch, the criminal category number corresponding to the case, and the set number. For the purposes of the study, several copies of the files exist. Each complete set of files has been given a "set" number for recordkeeping and accounting purposes.
2. Each large envelope contains two smaller envelopes containing the case file content. Envelope 1 contains the pre-polygraph case information. Envelope 2 contains the post-polygraph information. The polygraph questions and results have been removed or blackened out from the contents of Envelope 2. The files have been "sanitized" to remove actual names. Subjects have been identified as S1, S2, S3, etc.; witnesses as W1, W2, W3, etc.; observers as O1, O2, O3, etc.; and victims as V1, V2, V3, etc. On the outside of Envelope 1 and 2 are the case number, the envelope number, the criminal category number, and the set number.
3. A set of ratings forms are provided for each case. Each set contains the following forms:
  - a. Envelope 1: Global Rating 1,
  - b. Envelope 1: Initial Rating Form,
  - c. Rating sheet used when completing Initial Rating Form,
  - d. Envelope 2: Final Rating Form,
  - e. Rating sheet used when completing Final Rating Form, and
  - f. Envelope 2: Global Rating 2.4. A pad of paper and pen for taking notes while reviewing and evaluating cases.
5. Criminal category descriptions containing very brief descriptions of the relevant criminal categories.

Figure 5 continued

Case Review Process

**B. Procedure**

1. Open one large envelope, pull out the two smaller envelopes.
2. Open Envelope 1 and read the pre-polygraph case content. Use the note pad provided you to take notes. Include whatever information you feel necessary to review and evaluate the case. Keep your notes separate for each case and place the case number next to the notes written.
3. Complete Envelope 1: Global Rating 1 Form. If necessary, refer to the Criminal Category Description handout for a brief explanation of the criminal category.
4. Complete Envelope 1: Initial Rating Form. Use the rating sheet attached.
5. Return the contents back into Envelope 1 in the same order as you found them.
6. Open Envelope 2 and read the post-polygraph case content. Use your note pad for additional notes if necessary.
7. Complete Envelope 2: Final Rating form. Use the attached rating sheet.
8. Complete Envelope 2: global Rating 2 Form.
9. Return the contents back into Envelope 2 in the same order as you found them.
10. Return envelope 1 and 2 back into the large envelope.
11. continue steps 1-10 above until all cases are completed.
12. Return all materials, including note pads to the appropriate PDRI personnel.

## **Interrater Reliability of Panel Ratings**

In the behavioral sciences, when humans are required to make multiple ratings as did our panel members in reviewing case files, inconsistencies in the ratings are inevitable leading to what is generally referred to as measurement error. Measurement error can seriously affect statistical analyses and interpretation. The amount of measurement error is assessed by calculating a reliability index. Most measures of interrater reliability use variations of the formulae for computing the Intraclass Correlation Coefficient (ICC). The intraclass correlation is typically expressed as a ratio of the variance of interest over the variance of interest plus error. Depending upon how the error variance term is defined, several models for calculating ICC exist.

Shroat and Fleiss (1979) provide estimates of interrater reliability using various models for calculating the intraclass correlation coefficient. Typically, for interrater reliability studies, each of a random sample of  $n$  targets (e.g., global ratings for cases) is rated independently by  $k$  judges (e.g., panel members). Under the assumptions of Shroat and Fleiss' Model 2, a random sample of  $k$  judges is selected from a larger population, and each judge rates each target, that is, each judge rates  $n$  targets.

For the present study, Shroat and Fleiss' Model 2 was employed and measures of interrater reliability were calculated using the Average Composite Reliability. The Intraclass Correlation is a measure of the reliability of a single rater which could be considered a panel of one individual, although the calculation of this index must be performed on more than one rater at a time. The ICC for a single rater should be reported when future trials will use only a single rater or judge. The Average Composite Reliability is a measure of the reliability of a panel of raters and is the appropriate measure to report when a panel of more than one individual will be used to make the actual ratings in the future as in the case of a jury trial. The reliability calculations were based on all suspects that were rated. Since not all the suspects were from the same case, the reliabilities were corrected for case differences by removing the sum of squares and degrees of freedom associated with the case main effect. In other words, differences in ratings due to

case differences did not affect the reliability index. Both reliability indices were calculated using a two-way analysis of variance model. The formula used was derived from Shroat and Fleiss' Model 2, except that the variance due to each crime was removed from the between groups mean square (EMS).

The expected reliability of a single judges' ratings (i.e., single rater reliability or ICC) for Model 2 is calculated by using a Target and Judges two-way ANOVA. This analysis partitions the within-target sum of squares into a between judges mean square (JMS) and a residual sum of squares (EMS).

The computational formula is as follows:

$$ICC(2,1) = \frac{BMS - EMS}{BMS + (K-1)EMS + K(JMS-EMS)/n}$$

where n is the number of Targets (or Suspects); k is the number of raters, and BMS (Between Targets Mean Square) adapted as follows to:

$$BMS = (SS_{\text{between}} - SS_{\text{cases}}) / dF_{(\text{suspects-cases})}$$

Often it is not the individual ratings that are used but rather the mean of m ratings, where m need not be equal to k, the number of judges. The reliability of the mean rating (i.e., average composite reliability) will always be greater in magnitude than the reliability of the individual ratings (Shroat and Fleiss, p. 426). Typically, investigators use the mean rating because the individual rating is too unreliable.

The corresponding computational formula for the reliability of the ratings averaged over m judges (i.e., average composite reliability) is:

$$ICC(2,K) = \frac{BMS - EMS}{BMS + (JMS-EMS)/n}$$

### *Reliability of Specific Questions*

Table 22 presents the average composite reliability of the specific questions for each case rated by the three six-member panels. For all panels, the reliability of the final rating of the specific relevant questions is higher than the initial ratings of the same set of specific questions. The mean reliability of the initial ratings of the specific questions for the 27 cases, averaged across the three panels, is 0.182. For the final ratings of those same questions for the same groups, the reliability was .625

### *Reliability of Global Ratings*

Calculations for average composite reliability for the Global Ratings were done for panel sizes of 2, 3, 4, 5, and 6 within a rater group (i.e., lay person, police, and lawyer). Rater groups were never mixed together in a panel (i.e., lay persons were never combined with police or lawyers on a panel). Reliabilities were computed for all possible combinations of 6 raters taken n at a time where n was either 2, 3, 4, 5, or 6. For panels of 5 raters, this would yield six reliability estimates. For panels of 4 raters, this would yield 15 reliability estimates. The 15 reliability estimates constitutes the reliabilities of all possible panels of 4 that could be formed from six lawyers. If the lawyers were numbered 1 through 6, then the 15 panels would be:

(1,2,3,4), (1,2,3,5), (1,2,3,6), (1,2,4,5), (1,2,4,6), (1,2,5,6), (1,3,4,5), (1,3,4,6), (1,3,5,6),  
 (1,4,5,6), (2,3,4,5), (2,3,4,6), (2,3,5,6), (2,4,5,6), and (3,4,5,6)

Table 23 presents the number of reliability estimates or different panel arrangements for the corresponding panel size when there are six raters in all:

The means and standard deviations of the reliability estimates for the Global Ratings are reported in Tables 24 through 28. Table 24 shows the Average Composite Reliability that is expected for a panel of six individuals rating a suspect's guilt or innocence, and so on through Table 28

Table 22

Average Composite Reliability of Specific Questions  
(Panel Size of Six)

(Means for Each Case)<sup>a</sup>

CASE	Initial Rating (Specific Questions)					Final Rating (Specific Questions)				
	Lay	Police	Lawyer	Average of		Lay	Police	Lawyer	Average of	
	MEAN	MEAN	MEAN	MEAN	S.D.	MEAN	MEAN	MEAN	MEAN	S.D.
1.....	.082	.131	.291	.168	.28	.644	.710	.509	.621	.36
2.....	.118	.086	.070	.091	.24	.841	.740	.711	.764	.33
3.....	.735	.215	.000	.316	.35	.909	.925	.922	.918	.04
4.....	.138	.119	.253	.170	.30	.433	.406	.267	.369	.41
5.....	.517	.140	.737	.465	.38	.680	.805	.693	.726	.25
6.....	.000	.000	.640	.213	.32	.702	.726	.847	.758	.33
7.....	.207	.089	.159	.152	.29	.399	.173	.442	.338	.37
8.....	.000	.000	.213	.071	.18	.422	.859	.921	.734	.32
10....	.124	.049	.379	.184	.27	.517	.452	.399	.456	.30
11....	.089	.116	.075	.093	.18	.480	.773	.933	.729	.35
12....	.141	.219	.324	.228	.14	.591	.714	.669	.658	.20
13....	.256	.223	.318	.266	.27	.874	.854	.792	.840	.27
14....	.900	.533	.483	.639	.23	.735	.000	.800	.512	.44
15....	.232	.000	.511	.247	.35	.296	.459	.306	.354	.41
16....	.000	.199	.000	.066	.23	.885	.616	.932	.811	.30
17....	.353	.000	.736	.363	.35	.929	.357	.933	.740	.32
18....	.130	.000	.000	.043	.16	.829	.876	.736	.814	.28
19....	.074	.117	.043	.078	.20	.846	.939	.887	.891	.16
20....	.173	.327	.241	.247	.21	.467	.300	.349	.372	.32
21....	.119	.179	.306	.201	.31	.239	.110	.202	.184	.35
23....	.510	.077	.703	.430	.30	.817	.709	.680	.735	.36
24....	.303	.068	.409	.260	.37	.699	.709	.753	.720	.36
25....	.065	.044	.067	.058	.17	.588	.616	.620	.608	.40
26....	.167	.000	.000	.056	.18	.808	.678	.530	.672	.44
27....	.000	.135	.166	.100	.24	.397	.709	.816	.641	.31
Mean....	.178	.106	.261	.182	.29	.626	.617	.632	.625	.38

<sup>a</sup>Negative coefficients were converted to zero

Table 23

Number of Reliability Estimates by Panel Size

<u>Panel Size</u>	<u>Number of Estimates</u>
6	1
5	6
4	15
3	20
2	15

Table 24

Overall Reliabilities - for Global Ratings -Corrected for Case  
All Combinations of 6 at a Time

Average Composite Reliability		
	Mean	Std.Dev
<b>Global Rating 1</b>		
Layperson	.711	.
Police	.723	.
Lawyer	.806	.
Group Totals	.747	.052
<b>Global Rating 2</b>		
Layperson	.919	.
Police	.934	.
Lawyer	.924	.
Group Totals	.926	.008

Table 25

Overall Reliabilities - for Global Ratings Corrected for Case  
All Combinations of 5 at a Time

Average Composite Reliability		
	Mean	Std.Dev
<b>Global Rating 1</b>		
Layperson	.666	.054
Police	.686	.030
Lawyer	.768	.004
Group Totals	.707	.056
<b>Global Rating 2</b>		
Layperson	.903	.003
Police	.921	.011
Lawyer	.908	.008
Group Totals	.911	.011

Table 26

Overall Reliabilities - for Global Ratings Corrected for Case  
All Combinations of 4 at a Time

Average Composite Reliability		
	Mean	Std.Dev
<b>Global Rating 1</b>		
Layperson	.607	.075
Police	.618	.058
Lawyer	.730	.031
Group Totals	.651	.080
<b>Global Rating 2</b>		
Layperson	.883	.008
Police	.902	.018
Lawyer	.889	.015
Group Totals	.891	.016

Table 27

Overall Reliabilities - for Global Rating Corrected for Case  
All Combinations of 3 at a Time

Average Composite Reliability		
	Mean	Std.DEV
<b>Global Rating 1</b>		
Layperson	.524	.11
Police	.528	.10
Lawyer	.663	.06
Group Totals	.571	.11
<b>Global Rating 2</b>		
Layperson	.849	.01
Police	.871	.03
Lawyer	.856	.02
Group Totals	.858	.02

Table 28

Overall Reliabilities - for Global Ratings Corrected for Case  
All Combinations of 2 at a Time

Average Composite Reliability		
	Mean	Std.Dev
<b>Global Rating 1</b>		
Layperson	.401	.161
Police	.387	.193
Lawyer	.551	.126
Group Totals	.446	.175
<b>Global Rating 2</b>		
Layperson	.787	.041
Police	.812	.072
Lawyer	.795	.050
Group Totals	.798	.056

which shows the Average Composite Reliability that is expected for a panel of two individuals rating a suspect's guilt or innocence. "Group Totals" reported in the Tables are the mean and standard deviation of the reliability estimates across all three rater groups. For example, for groups of 4 raters at a time, it would be the mean and standard deviation of 45 reliability estimates (i.e., 15 reliability estimates from each of the three rater groups).

When all six members of each panel were included in the analysis (Table 24) the average composite Global 1 reliability is 0.71 for the lay panels, 0.72 for the police panel, and 0.81 for the attorney panel. The average Global 1 rating reliability across all groups is 0.75. The reliability of the average Global 2 rating is 0.92 for the lay panel, 0.93 for the police panel, and 0.92 for the attorney panel. The reliability of the average Global 2 rating across all groups is 0.93.

For all panels, regardless of size, (Tables 24-28) the Global Rating 2 data yielded higher reliability coefficients for each group than the Global Rating 1. This increase can be partly attributed to the fact that Global Rating 2 was made after reviewing additional case information, including the confessions of the guilty suspects.

The Average Composite Reliability estimates for the Global Ratings are graphically shown in Figures 6 through 9. Not surprisingly, it is clearly shown that the reliability estimates of the final (i.e., Global 2) ratings exceed the reliability estimates of the initial (i.e., Global 1) ratings regardless of panel size. Also, the larger the panel, i.e., the more people rating each case, the higher the overall or composite reliability.

It was surprising, however, that even when presented with the complete file, including the confession, the reliability of the panels was not near unity. This is an important finding because it indicates that confessions themselves, long considered an approximate ground truth criterion, are likely to be imperfect indicators of ground truth.

Figure 6

Average Composite Reliability of Initial and Final Ratings  
by Size of Layperson Panel

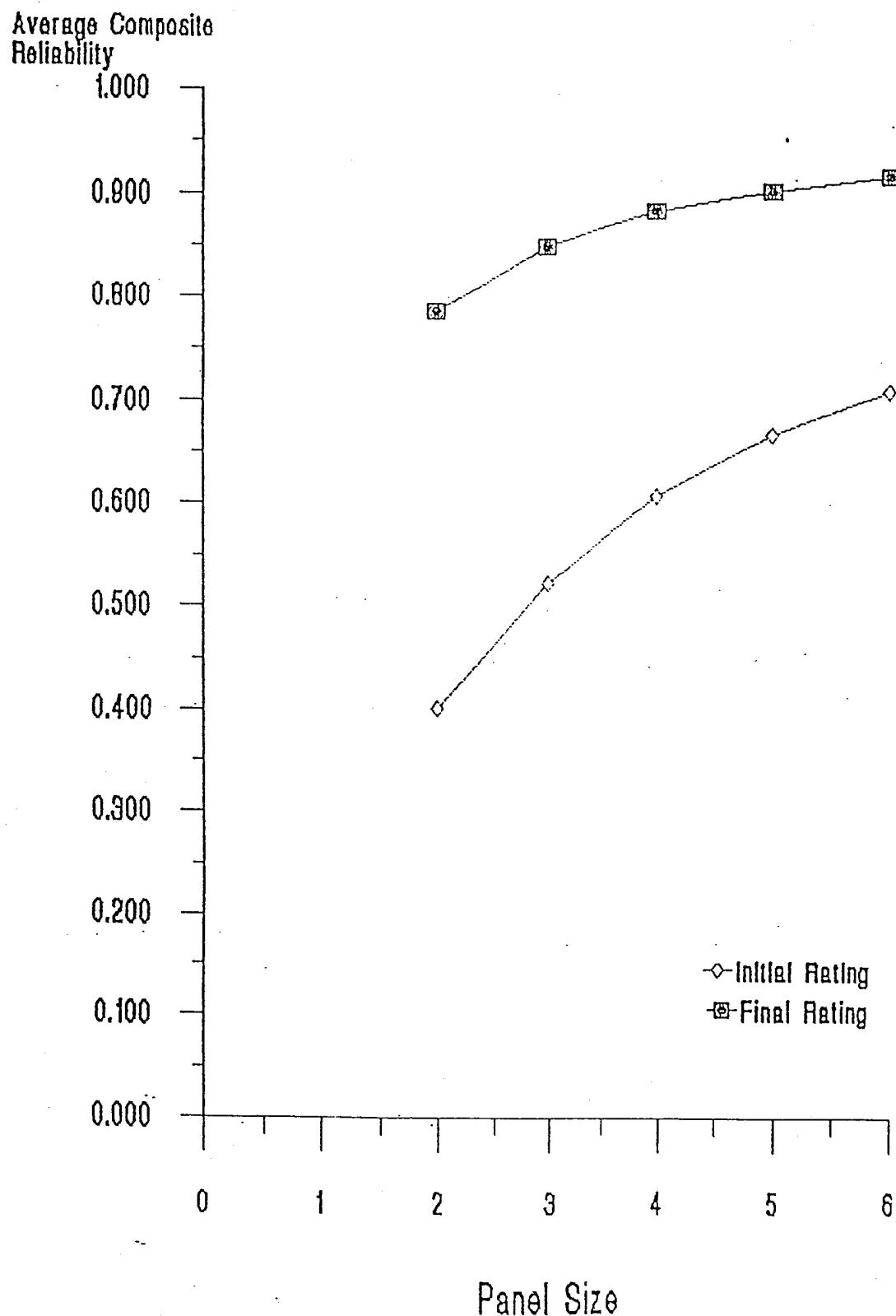
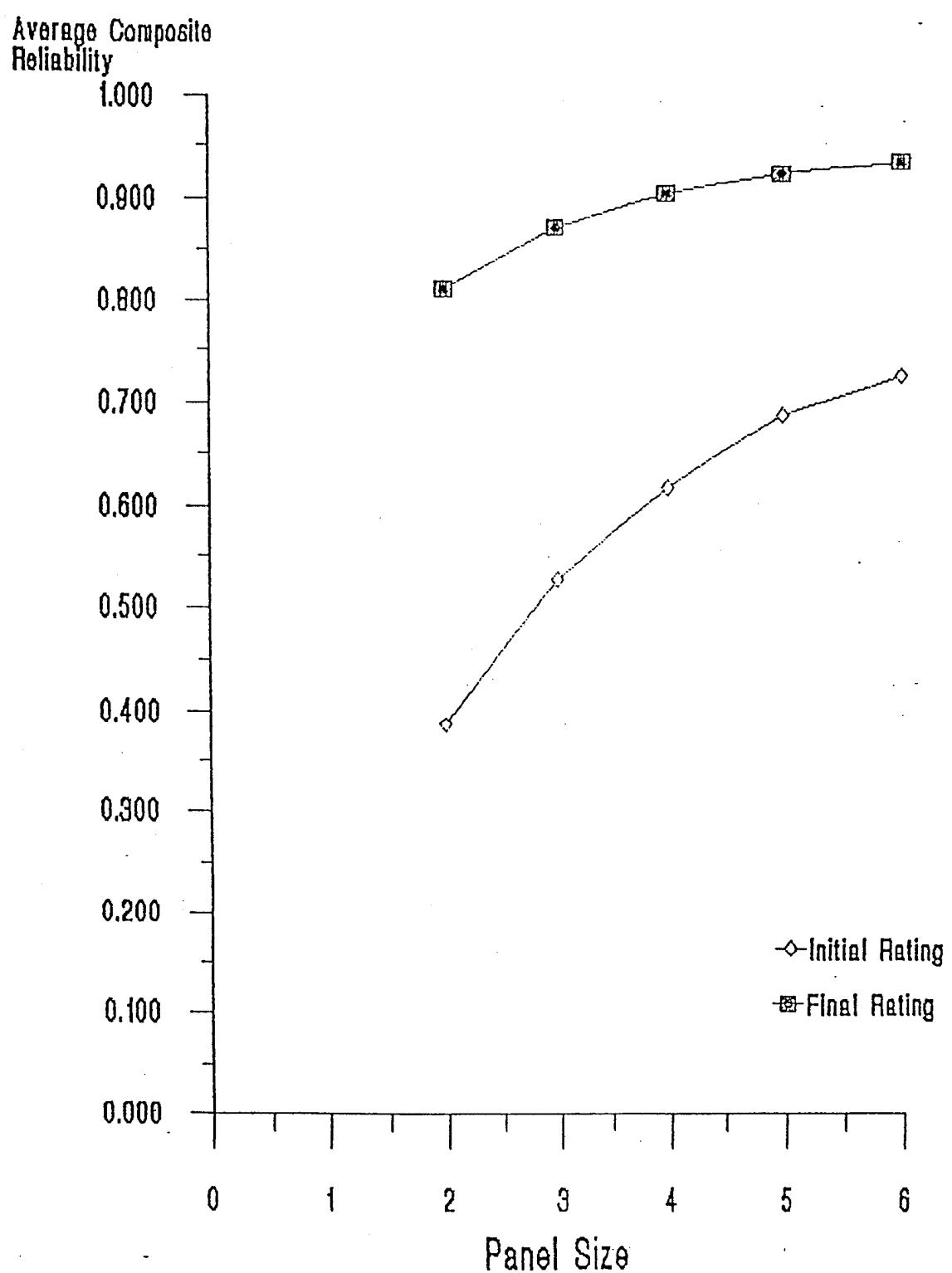


Figure 7  
Average Composite Reliability of Initial and Final Ratings  
by Size of Police Panel



**Figure 8**

**Average Composite Reliability of Initial and Final Ratings  
by Size of Lawyer Panel**

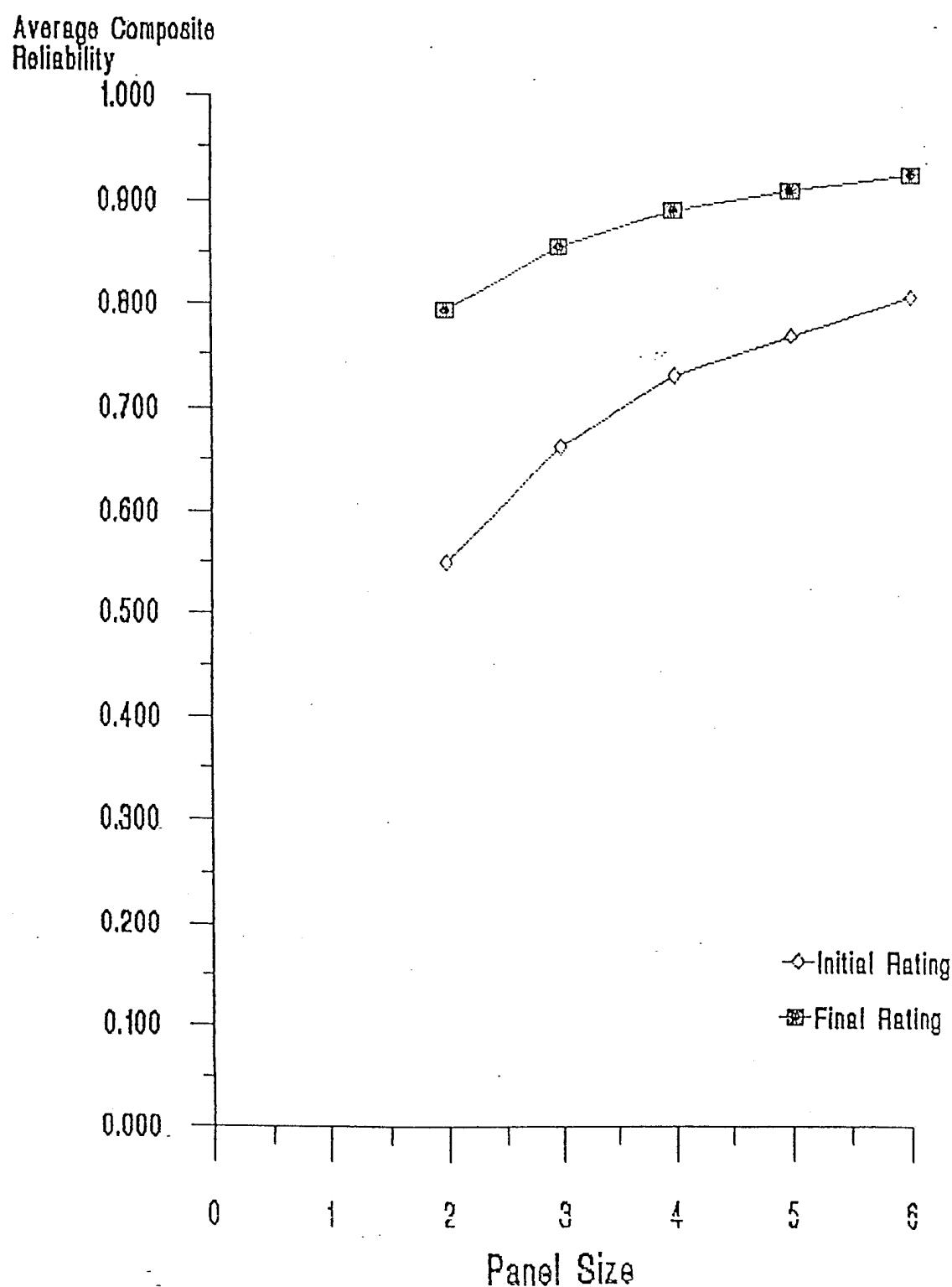
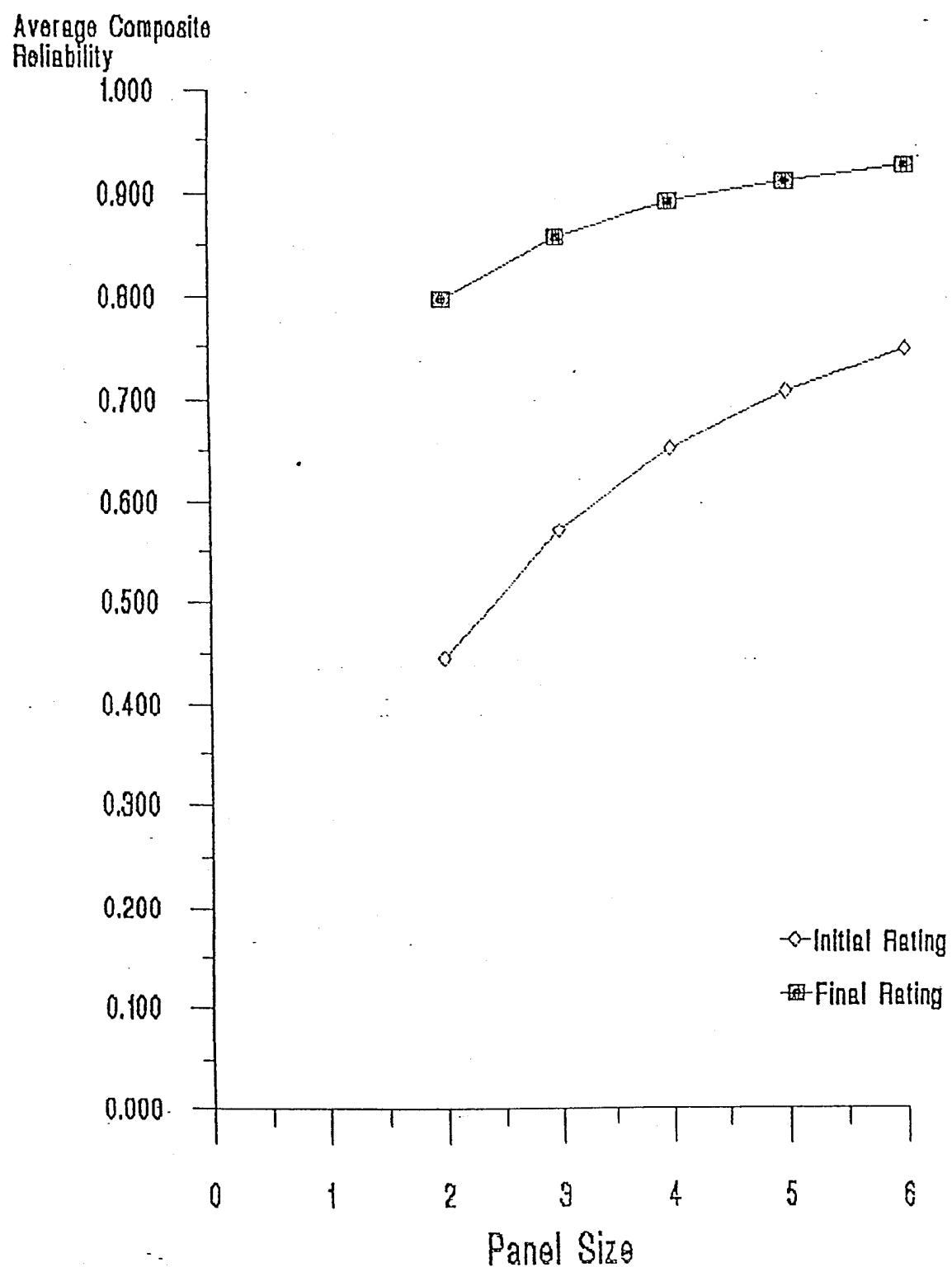


Figure 9

Average Composite Reliability of Initial and Final Ratings  
by Size of Combined Panel



If based on reliability estimates, the optimal panel size would vary depending upon whether or not partial or complete information is provided to the panel members. assuming we wish to maintain reliability of 0.70, then we would need to retain a panel of six members for the partial information (i.e., Global 1 Rating) condition. On the otherhand, if complete information provided to the panel, as had been done when the Global 2 ratings were made, a smaller panel size of even just two or three members would be sufficient to produce reasonably high reliability coefficients. As the panels increase in size from two to six there is only a marginal improvement in reliability.

### **Reliability of Specific Cases**

The previous section provided estimates of the overall reliability of the ratings. A closer examination of the average composite reliability of each case (using the six member panels) reveal differences among the various cases. Table 29 provides the average composite reliabilities of the Global Ratings by case. Overall, the lowest reliability estimates for Global 1 ratings were obtained for cases 4, 8, 15, 20, 26, and 27. The highest Global 1 reliability estimates were obtained for cases 5, 17, 25, 10, and 23 where the average reliability of the three panels ranged from 0.810 to 0.880. The lowest Global 2 reliability estimates were obtained for Cases 7, 4, 26, and 8. The highest Global 2 reliability estimates were found for cases 11, 14, 2, 5, 16, 23, 19, 12, 13, and 22 where the average reliability of the three panels ranged from 0.981 to 0.907.

The mean reliability of the Global 1 ratings across the 27 cases were 0.481 (Lay Panel), 0.497 (Police Panel), and 0.457 (Lawyer Panel). The mean reliability of the global 2 ratings across the 27 cases were 0.801 (Lay Panel), 0.758 (Police Panel), and 0.769 (Lawyer Panel).

Table 29

Average Composite Reliabilities of Global Ratings by Case<sup>a</sup>

All Combinations of 6 at a Time

## GLOBAL RATINGS

CASE	Layperson		Police		Lawyer		Average of Three Panels		Layperson		Police		Lawyer		Average of Three Panels	
	MEAN	SD.	MEAN	SD.	MEAN	SD.	MEAN	SD.	MEAN	SD.	MEAN	SD.	MEAN	SD.	MEAN	SD.
1	.194	.107	.942	.414	.46	.944	.817	.831	.864	.07	.964	.993	.964	.993	.964	.07
2	.649	.765	.000	.471	.41	.964	.935	.371	.717	.30	.000	.273	.091	.091	.091	.16
3	.000	1.000	1.000	.667	.58	.932	.848	.000	.273	.091	.000	.273	.091	.091	.091	.16
4	.000	.000	.000	.000	.00	.949	.957	.978	.961	.01	.000	.000	.000	.000	.000	.00
5	.864	.791	.986	.880	.10	.949	.957	.978	.961	.01	.000	.000	.000	.000	.000	.00
6	.600	.000	1.000	.533	.50	.996	.945	.729	.890	.14	.000	.000	.000	.000	.000	.00
7	.714	.819	.750	.761	.05	.000	.000	.000	.000	.00	.000	.000	.000	.000	.000	.00
8	.000	.000	.000	.000	.00	.794	.927	.965	.895	.09	.000	.000	.000	.000	.000	.09
9	.000	.187	.767	.318	.40	.741	.000	.000	.000	.00	.000	.000	.000	.000	.000	.43
10	.848	.832	.774	.818	.04	.720	.603	.547	.623	.09	.000	.000	.000	.000	.000	.09
11	1.000	1.000	.000	.667	.58	.987	.996	.955	.979	.02	.000	.000	.000	.000	.000	.02
12	.775	.477	.924	.725	.23	.952	.891	.932	.925	.03	.000	.000	.000	.000	.000	.03
13	.000	.632	.000	.211	.36	.930	.863	.968	.920	.05	.000	.000	.000	.000	.000	.05
14	.516	.581	.000	.366	.32	.992	.962	.984	.979	.02	.000	.000	.000	.000	.000	.02
15	.000	.000	.000	.000	.00	.881	.914	.829	.875	.04	.000	.000	.000	.000	.000	.04
16	.822	.000	.286	.369	.42	.937	.993	.992	.974	.03	.000	.000	.000	.000	.000	.03
17	.812	.914	.906	.877	.06	.936	.992	.956	.961	.03	.000	.000	.000	.000	.000	.03
18	.649	1.000	.600	.750	.22	.960	.997	.987	.981	.02	.000	.000	.000	.000	.000	.02
19	.667	.862	.774	.768	.10	.891	.966	.934	.930	.04	.000	.000	.000	.000	.000	.04
20	.000	.000	.000	.000	.00	.809	.927	.833	.856	.06	.000	.000	.000	.000	.000	.06
21	.700	.898	.800	.799	.10	.856	.934	.930	.907	.04	.000	.000	.000	.000	.000	.04
22	.689	.000	.000	.230	.40	.784	.927	.669	.793	.13	.000	.000	.000	.000	.000	.13
23	1.000	1.000	.429	.810	.33	.983	.973	.910	.955	.04	.000	.000	.000	.000	.000	.04
24	.706	.667	.533	.635	.09	.764	.830	.802	.799	.03	.000	.000	.000	.000	.000	.03
25	.789	.889	.864	.847	.05	.653	.840	.840	.778	.11	.000	.000	.000	.000	.000	.11
26	.000	.000	.000	.000	.00	.450	.000	.000	.000	.26	.000	.000	.000	.000	.000	.26
27	.000	.000	.000	.000	.00	.831	.429	.988	.749	.29	.000	.000	.000	.000	.000	.29

Mean .481

.497

.457

.478

.40

.801

.758

.748

.769

.31

<sup>a</sup>Negative coefficients were converted to zero.

For the large majority of cases, the Global 2 reliability estimates were higher than the Global 1 estimates. In a few cases however, most notably in Case 7 and to a lesser extent in Cases 9, 10 and 24, the introduction of additional case information had produced less reliable Global 2 than Global 1 ratings.

In terms of cost considerations for the present study, the attorney panel members were paid 3.87 times more than the lay panel members and 2.69 times more than the police panel members. The police panel members were paid 1.44 times more than the lay panel. The differences in reliability between lay, police and lawyers on the final ratings are so small given the huge differences in cost for these 3 types of raters, a cost/benefit analysis would recommend using lay people as raters if reliability is the only consideration.

### **Global Rating Changes**

The Global ratings were made on a five point scale:

- 1 - Definitely did not commit the above crime**
- 2 - Probably did not commit the above crime**
- 3 - Undecided**
- 4 - Probably did commit the above crime**
- 5 - Definitely did commit the above crime**

The mean Global ratings for each panel for each suspect/victim in each of the 27 cases is provided in Table 30. The "Combined" rating is the average of the ratings of the 18 panel members (6 Lay, 6 Police, 6 Lawyers).

Table 30<sup>a</sup>  
Mean Global Ratings by Suspect/victim

GLOBAL I RATINGS

CASE	LAWYER		POLICE		LAWYER		POLICE		LAWYER		GLOBAL 2 RATINGS		
	MEAN	S.D.	MEAN	S.D.									
1	S1	3.00	.63	2.50	.55	2.17	.41	2.56	.62	2.17	.98	2.33	1.51
	S2	3.33	.52	3.17	.98	3.50	.55	3.33	.69	4.50	.55	3.50	1.64
	O1	2.83	.41	3.00	.63	3.00	.00	2.94	.42	2.17	.75	1.00	.00
2	S1	3.50	.55	4.00	.63	3.17	.75	3.56	.70	4.83	.41	4.83	.41
	S2	2.83	.75	3.00	.10	3.33	.52	3.06	.80	2.17	.17	2.17	1.60
	CS1	3.00	.00	3.00	.00	3.00	.00	3.00	.00	3.00	.10	3.50	.55
3	CS2	2.83	.41	3.00	.00	3.00	.00	2.94	.24	4.83	.41	4.50	.84
	V1/S	3.50	.55	3.50	.84	3.33	.52	3.44	.62	4.83	.41	4.83	.41
	S1	3.33	.52	3.50	.84	3.50	.55	3.44	.62	4.83	.41	4.83	.41
4	S1	4.67	.52	4.17	.75	4.50	.55	4.44	.62	5.00	.00	5.00	.00
	S2	3.00	.41	2.67	.37	2.17	.41	2.61	.14	2.50	.38	1.83	1.60
5	S1	4.67	.52	4.17	.75	4.50	.55	4.44	.62	5.00	.00	5.00	.00
	S2	3.00	.41	2.67	.37	2.17	.41	2.61	.14	2.50	.38	1.83	1.60

Global Rating Scale:

- 1 = Definitely did not commit the above crime
- 2 = Probably did not commit the above crime
- 3 = Undecided
- 4 = Probably did commit the above crime
- 5 = Definitely did commit the above crime

Table 30 continued

### Mean Global Ratings by Suspect/victim

GLOBAL I RATINGS

LAWYER		POLICE		LAWYER		COMBINED		LAY		POLICE		LAWYER		COMBINED	
MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
6															
S2	.3.33	.52	3.17	.41	3.00	.00	3.17	.38	1.50	.55	1.33	.52	2.00	1.67	1.61
S1	3.00	.00	3.00	.00	3.00	.00	3.00	.00	5.00	.00	4.33	1.63	4.00	1.55	4.44
7															
S1c1	2.83	.41	2.67	.82	3.00	.63	2.83	.62	4.17	.17	4.50	.55	4.00	1.26	4.22
S2c1	2.67	.82	2.67	.82	3.00	.00	2.78	.65	4.17	.17	4.33	.52	4.00	1.10	4.17
S1c2	3.67	.52	4.17	.75	3.67	.52	3.83	.62	3.33	.63	4.33	1.21	3.67	1.21	3.78
S2c2	2.67	.82	2.67	.82	3.00	.00	2.78	.65	4.17	.17	4.50	.55	4.33	.52	4.33
3															
S2	3.50	.84	3.17	.75	3.67	.52	3.44	.70	4.67	.52	4.83	.41	4.67	.52	4.72
S3	3.50	.55	3.17	.75	3.50	.55	3.39	.61	3.50	.22	2.67	1.37	2.00	1.10	2.72
3															
S1	4.17	.33	3.67	.03	2.67	.03	3.50	.25	3.33	.86	4.33	1.21	3.50	1.64	3.72
S2	4.33	.52	3.33	.82	3.67	.52	3.78	.73	4.83	.41	4.17	.75	4.00	1.10	4.33

Table 30 continued

## Mean Global Ratings by Suspect/Victim

## GLOBAL 1 RATINGS

	GLOBAL 1 RATINGS						GLOBAL 2 RATINGS					
	LAWYER		POLICE		COMBINED		LAWYER		POLICE		COMBINED	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
10												
S1	4.00	.00	4.00	.63	3.67	.52	3.89	.47	4.83	.41	5.00	.00
S5	2.50	.84	3.17	.41	3.00	.00	2.89	.58	3.17	.33	3.33	1.86
S4	2.50	.84	3.00	.00	3.00	.00	2.83	.51	3.33	.51	4.17	1.17
S3	2.50	.84	3.33	.52	2.67	.82	2.83	.79	4.50	.55	4.17	.41
11												
S2	3.00	.00	3.00	.00	2.83	.41	2.94	.24	1.67	.52	1.17	.41
S1	3.00	.00	3.00	.00	3.00	.00	3.00	.00	4.33	.52	4.83	.41
12												
V1	2.50	.55	2.50	.55	2.50	.55	2.50	.51	1.50	.55	1.67	1.21
S2	2.33	.21	2.83	.98	2.00	.63	2.39	.98	2.00	.26	2.50	1.76
S1	3.67	.52	3.33	.52	3.67	.52	3.56	.51	4.50	.55	4.67	.52
13												
S2	3.17	.41	3.00	.00	3.33	.52	3.17	.38	2.00	.26	2.33	1.37
S5	3.00	.00	3.50	.55	3.17	.41	3.22	.43	2.00	.26	2.50	1.64
S1	3.00	.00	3.17	.41	3.33	.52	3.17	.38	4.83	.41	4.83	.41

Table 30 continued

## Mean Global Ratings by Suspect/Victim

## GLOBAL 1 RATINGS

	GLOBAL 1 RATINGS						GLOBAL 2 RATINGS					
	LAWYER		POLICE		COMBINED		LAWYER		POLICE		LAWYER	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
14												
S1	3.33	.82	2.83	.41	3.00	.89	3.06	.73	1.33	.52	1.83	1.17
S2c	2.50	.84	3.00	.00	2.83	.41	2.78	.55	4.83	.41	4.83	.67
S2c	2.50	.84	3.33	.52	2.83	.41	2.89	.68	4.83	.41	4.50	.55
15												
S1	3.00	.00	3.00	.00	2.50	.55	2.83	.38	3.67	.03	3.83	1.17
S2	2.67	.03	2.83	.41	2.50	.84	2.67	.77	1.83	.17	2.00	.63
16												
V2c1	3.50	.55	2.83	.75	3.33	.82	3.22	.73	1.17	.41	1.67	.52
S1c2	2.50	.55	3.17	.41	2.67	.52	2.78	.55	5.00	.00	5.00	.00
V2c1	3.33	.52	3.00	.63	2.83	.75	3.06	.64	2.33	.07	1.33	.52
17												
V1c2	2.50	.55	2.50	.55	1.83	.41	2.28	.57	1.50	.84	1.17	.41
V1c2	2.50	.55	2.50	.55	1.83	.41	2.28	.57	1.50	.84	1.17	.41
S1	3.50	.55	3.83	.41	3.67	.52	3.67	.49	4.83	.41	4.83	.41
S2	2.50	.55	2.33	.52	2.33	.03	2.39	.70	2.00	.55	1.17	.41

Table 30 continued

## Mean Global Ratings by Suspect/Victim

GLOBAL 1 RATINGS												GLOBAL 2 RATINGS												
	LAWYER						POLICE						LAWYER						POLICE					
	LAY		POLICE		LAWYER		LAY		POLICE		LAWYER		LAY		POLICE		LAWYER		POLICE		LAWYER		COMBINED	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
18																								
	S2	.2.67	.52	3.00	.00	3.33	.52	3.00	.49	1.83	.98	1.33	.52	1.67	.52	1.61	1.61							
	S1	3.17	.41	3.00	.00	3.00	.00	3.06	.24	4.50	.84	5.00	.00	4.33	.52	4.61	4.61							
19																								
	S3	3.33	.52	3.83	.41	3.50	.55	3.56	.51	2.17	.47	2.33	.82	2.33	1.51	2.28	2.28							
	S1	2.83	.41	2.83	.41	2.83	.009	.009	.32	.003	.003	.41	.41	.41	.41	4.89	4.89							
	S2	2.83	.41	3.00	.63	2.83	.41	2.89	.47	3.33	.37	2.17	.75	2.00	.63	2.50	2.50							
20																								
	S1	3.17	.41	3.33	.52	3.33	.82	3.28	.57	2.83	.98	2.50	.55	2.50	1.05	2.61	2.61							
	S4	3.17	.41	3.33	.52	3.50	.55	3.33	.49	3.83	.41	3.67	.52	3.67	.52	3.72	3.72							
21																								
	S1c1	3.33	.82	2.83	.41	3.50	.05	3.22	.81	2.83	.17	2.33	.82	2.33	.82	2.50	2.50							
	S2c1	2.67	.82	2.83	.41	2.67	.52	2.72	.57	2.17	.98	2.17	.75	2.00	.00	2.11	2.11							
	S1c2	4.00	.89	4.33	.52	4.17	.98	4.17	.79	4.67	.52	4.83	.41	4.33	1.21	4.61	4.61							
	S2c2	2.67	.82	3.33	.82	2.83	.41	2.94	.73	2.50	.38	2.17	.17	1.17	2.00	.00	2.22	2.22						

Table 30 continued

## Mean Global Ratings by Suspect/Victim

## GLOBAL I RATINGS

	GLOBAL I RATINGS						GLOBAL 2 RATINGS						POLICE					
	LAY			POLICE			LAWYER			POLICE			LAWYER			COMBINED		
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
22																		
S1c1	2.83	.41	3.00	.63	3.00	.63	2.94	.54	3.17	.60	1.67	.52	2.50	.64	2.44	2.44		
S1c2	3.17	.41	3.00	.63	2.83	.41	3.00	.49	4.50	.55	4.00	1.55	3.67	1.37	4.06	4.06		
S2	3.50	.55	3.17	.41	3.17	.41	3.28	.46	4.83	.41	4.50	.55	4.33	.82	4.56	4.56		
23																		
S2c1	3.00	.00	3.00	.00	2.67	.52	2.89	.32	1.00	.00	1.00	.00	1.00	.00	1.00	1.00		
S1	3.00	.00	3.00	.00	3.00	.00	3.00	.00	5.00	.00	4.83	.41	4.00	1.26	4.61	4.61		
S2c2	3.00	.00	3.00	.00	2.67	.52	2.89	.32	1.83	.17	3.00	1.26	2.17	1.47	2.33	2.33		
24																		
S3	3.33	.52	2.83	.41	2.67	.52	2.94	.54	1.33	.82	1.50	.84	1.00	.00	1.28	1.28		
S2	3.17	.41	3.33	.52	3.17	.41	3.22	.43	3.50	.52	4.33	1.63	3.33	1.97	3.72	3.72		
S1	2.33	.03	2.67	.52	2.67	.52	2.56	.70	3.50	.97	2.67	1.63	1.67	1.21	2.61	2.61		
25																		
S1	3.33	.52	3.50	.55	3.50	.55	3.44	.51	4.83	.41	5.00	.00	5.00	.00	4.94	4.94		
S2	2.67	.52	2.50	.55	2.67	.52	2.61	.50	3.83	.47	3.33	1.63	3.33	1.63	3.50	3.50		
26																		
S1	2.83	.41	2.83	.41	3.00	.63	2.89	.47	4.33	.21	4.50	1.22	4.83	.41	4.56	4.56		
S2	3.17	.41	3.17	.41	3.00	.63	3.11	.47	5.00	.00	4.83	.41	4.83	.41	4.89	4.89		

Table 30 continued

## Mean Global Ratings by Suspect/Victim

## GLOBAL I RATINGS

## GLOBAL 2 RATINGS

	GLOBAL I RATINGS						GLOBAL 2 RATINGS					
	LAY		POLICE		LAWYER		LAY		POLICE		LAWYER	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
27												
S1	3.00	.63	3.33	.82	3.17	.75	3.17	.71	4.67	.82	4.67	.52
S2	3.00	.63	3.33	.82	3.00	.63	3.11	.68	3.17	.17	4.33	.52

Overall, did the global ratings change (i.e., the probability statement regarding whether or not subjects committed the crimes) for the various panels when panelists were given post-polygraph data that included confessions? As expected, they did indeed. Comparing the initial (Global 1) ratings with the final (Global 2) ratings across all cases, the ratings of each panel changed as a result of providing additional information. Significant chi-square statistics were obtained at the  $p < .001$  level for the lay panel (Table 31), the police officer panel (Table 32), and the attorney panel (Table 33). For all panels, the additional information reduced the amount of uncertainty in deciding whether or not a subject committed a crime. The number of "undecided" ratings decreased from 62.9% for the initial ratings to 8.9% for the final ratings made by the lay panel. The "undecided" ratings reduced from 63.6% to 7.5% for the police panel and from 60.1% to 8.7% for the attorney panel. From the data provided in Tables 31 to 33 it is clear that the initial data available prior to the polygraph exam is insufficient for most panel members to make a Global rating of guilt/innocence since more than 60% of the ratings of each panel was "Undecided". Again, it is important to note that even with confessions in hand, a small fraction of panelists were uncertain about the guilt or innocence of some suspects.

### **Leniency of Ratings**

Do the panels differ in terms of their global ratings? Are some panels more lenient (i.e., less likely to rate suspects as committing crimes) than other panels?

A series of one-way ANOVAs were run using the global ratings as the dependent variable and type of panel as the independent variables.

When the *initial* global rating (pre-polygraph information) was used as the dependent variable, no differences between panels was found ( $F=0.6898$ ,  $p=0.5169$ ).

Table 31  
Initial Versus Final Global Ratings  
(Lay Panel)

**Final Rating Panel**

Initial Rating	Definitely Did Not	Probably Did Not	Un-decided	Probably Did	Definitely Did	Row Total/ Percent
Definitely Did Not	7	1	0	1	3	12 2.8%
Probably Did Not	16	6	2	7	21	52 12.2%
Undecided	48	47	32	49	92	268 62.9%
Probably Did	12	6	4	21	38	81 19.0%
Definitely Did	1	0	0	1	11	13 3.1%
Column Total	84	60	38	79	165	426
Column Percent	19.7%	14.1%	8.9%	18.5%	38.7%	100.0%

Chi-Square (16df) = 44.82; p<.001

Table 32  
Initial Versus Final Global Ratings  
(Police Panel)

Final Rating Panel

Initial Rating	Definitely Did Not	Probably Did Not	Un-decided	Probably Did	Definitely Did	Row Total/ Percent
Definitely Did Not	3	0	0	3	0	6 1.4%
Probably Did Not	26	7	0	6	9	48 11.3%
Undecided	54	45	25	52	95	271 63.6%
Probably Did	9	8	7	15	52	91 21.4%
Definitely Did	1	0	0	0	9	10 2.3%
Column Total	93	60	32	76	165	426
Column Percent	21.8%	14.1%	7.5%	17.8%	38.7%	100.0%

Chi-Square (16df) = 72.74; p<.001

Table 33  
Initial Versus Final Global Ratings  
(Lawyer Panel)

Final Rating Panel

Initial Rating	Definitely Did Not	Probably Did Not	Un-decided	Probably Did	Definitely Did	Row Total/ Percent
Definitely Did Not	4	1	0	0	0	5 1.2%
Probably Did Not	28	25	3	7	12	75 17.6%
Undecided	39	61	28	55	73	256 60.1%
Probably Did	14	8	6	22	33	83 19.5%
Definitely Did	0	1	0	0	6	7 1.6%
Column Total	85	96	37	84	126	426
Column Percent	20.0%	22.5%	8.7%	19.7%	29.1%	100.0%

Chi-Square (16df) = 65.76; p<.001

When the final global rating (pre- and post-polygraph information) was used as the dependent variable, a significant mean difference was found ( $F=5.7262$ ,  $p=.0142$ ). Using the Multiple Range Test and LSD procedure, two homogeneous subgroups were formed. The first subgroup contained attorneys, the second subgroup contained lay persons and police officers. No significant mean difference in final global rating was found between the lay panel and the police panel. A significant mean difference in final global rating was found between the attorneys and the other two groups. The attorneys were more likely to rate the suspects as *not* committing the crime than the other two groups. (Mean rating for lawyers = 3.22, police = 3.43, and lay persons = 3.49).

### **Relationship Between Specific Questions to Global Ratings**

Panel members provided specific ratings to relevant questions as well as global ratings. The question thus arises as to which specific questions employed in the polygraph examination(s) significantly correlated with the global ratings. The responses of the 18 panel members to the specific questions on the Initial Rating Form were correlated with the responses to the Global Rating 1 Form. Likewise, the responses to the specific questions on the Final Rating Form were correlated with the responses to the Global Rating 2 Form. The results are presented in Appendix T. This set of analyses helps explain which specific items (asked during the polygraph exam) contribute most to explaining the variation in global ratings (one's belief as to whether or not the subject committed the crime). Generally, the number of specific items significantly correlating with global ratings was greater for the final Global 2 ratings, as compared with the Initial Global 1 ratings.

### **Accuracy of Outcome**

To obtain some measure of outcome accuracy, three types of information regarding subjects' guilt were collected and compared:

1. The suspect's polygraph results (i.e.,

DI=Deception Indicated

NDI=No Deception Indicated

INC=Inconclusive

2. Panel member's Global rating of suspects guilt, and
3. Independent judges' (i.e., principal investigators') rating as to whether or not suspect's confession of guilt was verified. Recall that, for the cases given to the panels, both independent judges agreed with each other on the suspect's guilt based on the Quality of the Confession (i.e., "The confession verified the accusation contained in the relevant question with almost certainty").

Accuracy of outcome was measured by comparing the panel members' ratings with the independent judges' ratings of ground truth. Ground truth is defined, in this instance, as the agreement by both independent judges that the case is "confession-verified" based on the quality of confession criteria mentioned above. In this case, a "hit" is defined when the independent judges agreed that the confession of guilt was verified *and* the panel member's global rating response indicated the suspect "probably did commit the above crime" or "definitely did commit the above crime." Thus, in this scenario, the hit rate is defined in terms of correctly identifying a "guilty" suspect. A "hit" could also occur if the confession of one suspect "cleared" or demonstrated the innocence of another suspect. For example, in a rape case, if a male suspect admits that he raped a female victim, then the victim's statement of "innocence" was verified. In this example, the independent judges indicated the confession demonstrated the victim's innocence and the panel member's rating indicated the victim "probably did not commit the above crime" or "definitely did not commit the above crime." Thus, hit rate is defined in terms of correctly identifying an "innocent" suspect. An "undecided" panel rating was considered a "miss." Likewise, a "miss" also occurred if a panel member indicated the suspect "probably did not commit the above crime" or "definitely did not commit the above crime" and the independent judges indicated the confession of guilt by the suspect was verified. Finally, a "miss" occurred if a panel member indicated the suspect "probably did commit the above crime" or "definitely did commit the above crime" and the independent judges indicated the confession verified the suspect's innocence.

Since it is fairly rare to obtain confession-verified innocence, the large majority of confession-verified cases in this study correctly identify guilty suspects. Thus, the hit rates reported in this study are, unless otherwise stated, overall hit rates, combining the correct "hits" of innocence and the correct "hits" of guilt. Hit rates are expressed as a proportion of percentage of the total.

The accuracy of outcome data can be presented in several ways:

1. What is the overall accuracy of the ratings? (i.e., the accuracy of the overall global ratings across all cases and all panel types).
2. What is the overall accuracy of the ratings *for each panel*? (i.e., the accuracy of the overall global ratings across all cases for each panel type).
3. What is the overall accuracy of the ratings *for each suspect*? (i.e., the accuracy of the overall global ratings for each suspect/victim in each case across all panel types).
4. What is the overall accuracy of the ratings *for each suspect for each panel*? (i.e., the accuracy of the overall global ratings for each panel type for each suspect/victim in each case).
5. What is the overall accuracy of the ratings *for each panel member*? (i.e., the accuracy of the overall global ratings for each panel member).
6. What is the overall accuracy of the *specific* ratings? (i.e., the accuracy of the specific relevant question ratings for each suspect/victim in each case).

#### **What is the Overall Accuracy of the Ratings?**

Table 34 provides the accuracy of the overall global ratings across all cases and all panel types. The accuracy of Global 1 Ratings is 22.5% for the combined panel of 18 members and 71.5%

Table 34  
Accuracy of Outcome: Global Ratings of Each Panel Across Cases<sup>a</sup>

Panel Type

Global Rating	Lay	Police	Lawyer	Combined
Global Rating 1 (Initial)	21.3% (U-2, U-3)	26.7% (U-6,U-7)	19.6% (U-10,U-11)	22.5% (U-14,U-15)
Global Rating 2 (Final)	72.9% (U-4,U-5)	74.2% (U-8,U-9)	67.5% (U-12,U-13)	71.5% (U-16,U-17)

<sup>a</sup> For more detailed information please refer to the appropriate appendices enclosed in parentheses.

for Global 2 Ratings. Appendix U provides additional details of the overall accuracy. The low hit rate for the Global 1 Rating is to a large extent due to the large percentage of "undecided" ratings made by panel members. Recall that undecided ratings are counted as "misses."

Due to the low hit rate, it would appear inadvisable to use panels to establish ground truth based on pre-polygraph test case data.

#### **What is the Overall Accuracy of the Ratings for Each Panel?**

Table 34 also provides data as to the accuracy of the overall Global Ratings across all cases for each panel type. The Global 2 "hit rates" were substantially larger than the Global 1 hit rates for all panels. The Global 2 hit rates as compared to the Global 1 hit rates were 72.9% to 21.3% (lay panel), 74.2% to 26.7% (police panel), and 67.5% to 19.6% (lawyer panel). The fact that the Global 2 "hit rates" were not 100% indicates that the panels did not consider the post-polygraph confessions to be satisfactory indices of ground truth about 35-33% of the time. Additional details regarding the accuracy of ratings for each panel may be found in Appendix U.

Once again, regardless of type of panel, the hit rates of the Global 1 ratings were quite low, indicating that panels, regardless of type should not be used to establish ground truth at an early stage of the investigation (i.e., prior to the administration of the polygraph exam).

#### **What is the Overall Accuracy of the Ratings for Each Suspect?**

Table 35 presents the accuracy of the overall global ratings for each suspect/victim in each case across all panel types. The accuracy of the Global 1 and Global 2 ratings are provided in the "combined" columns which is the mean hit rate of the 18 panel members for each suspect/victim. Confession data was not collected or obtained from every suspect in the case so for this subset of suspects an N/A ("Not Applicable") was placed in the column to indicate that outcome accuracy could not be determined.

Table 35

Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim  
of Each Case

	GLOBAL I (INITIAL) RATINGS				GLOBAL 2 (FINAL) RATINGS				
	Lay	Police	Lawyer	Combined	Lay	Police	Lawyer	Combined	
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	
<b>CASE</b>									
1									
	S1....	.17	.00	.00	.06	.17	.17	.00	.11
	S2....	.33	.50	.50	.44	1.00	.67	.83	.83
	O1....	.00	.17	.00	.06	.00	.00	.17	.06
2									
	S1....	.50	.83	.33	.56	1.00	1.00	1.00	1.00
	S2....	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3									
	CS1...	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	CS2...	.00	.00	.00	.00	1.00	.83	.83	.89
4									
	V1/S2.	.50	.67	.33	.50	1.00	1.00	.67	.89
	S1....	.00	.17	.00	.06	.00	.00	.00	.00
5									
	S1....	.00 <sup>a</sup>	.00	.00	.00	.00	.00	.00	.00
	S2....	.33	.33	.00	.22	.33	.17	.17	.22
6									
	S2....	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1....	.00	.00	.00	.00	1.00	.83	.83	.89
7									
	S1c1..	.00	.00	.17	.06	.83	1.00	.67	.83
	S2c1..	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1c2..	.67	.83	.67	.72	.50	.83	.50	.61
	S2c2..	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

<sup>a</sup>Not applicable: Confession data was not collected from this individual so outcome accuracy could not be determined.

Table 35 continued

Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim  
of Each Case

	GLOBAL I (INITIAL) RATINGS				GLOBAL 2 (FINAL) RATINGS			
	Lay	Police	Lawyer	Combined	Lay	Police	Lawyer	Combined
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN
8								
	S2....	.67	.33	.67	.56	1.00	1.00	1.00
	S3....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
9								
	S1....	.67	.67	.33	.56	.50	.83	.67
	S2....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
10								
	S1....	1.00	.83	.67	.83	1.00	1.00	.67
	S5....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S4....	.00	.00	.00	.00	.67	.83	.50
	S3....	.17	.33	.17	.22	1.00	1.00	.67
								.89
11								
	S2....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1....	.00	.00	.00	.00	1.00	1.00	1.00
12								
	V1....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S2....	.17	.17	.00	.11	.17	.33	.00
	S1....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
13								
	S2....	.00	.00	.00	.00	.67	.67	.83
	S5....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1....	.00	.17	.33	.17	1.00	1.00	1.00
14								
	S1....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S2c1..	.17	.00	.00	.06	1.00	1.00	1.00
	S2c2..	.17	.33	.00	.17	1.00	1.00	.83
								.94

Table 35 continued

Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim  
of Each Case

	GLOBAL I (INITIAL) RATINGS				GLOBAL 2 (FINAL) RATINGS			
	Lay	Police	Lawyer	Combined	Lay	Police	Lawyer	Combined
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN
15								
	S1.... .00	.00	.00	.00	.67	.67	.50	.61
	S2.... N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
16								
	V2c1.. N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1c2.. .00	.17	.00	.06	1.00	1.00	1.00	1.00
	V2c2.. N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
17								
	V1c1.. N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	V1c2.. N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1.... .50	.83	.67	.67	1.00	1.00	1.00	1.00
	S2.... N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
18								
	S2.... N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1.... .17	.00	.00	.06	.83	1.00	1.00	.94
19								
	S3.... .00	.00	.00	.00	.67	.50	.50	.56
	S1.... .00	.00	.00	.00	1.00	1.00	1.00	1.00
	S2.... .17	.17	.17	.17	.33	.67	.83	.61
20								
	S1.... .17	.33	.50	.33	.33	.00	.17	.17
	S4.... N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 35 continued

Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim  
of Each Case

	GLOBAL I (INITIAL) RATINGS				GLOBAL 2 (FINAL) RATINGS			
	Lay	Police	Lawyer	Combined	Lay	Police	Lawyer	Combined
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN
21								
	S1c1..	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S2c1..	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1c2..	.67	1.00	.67	.78	1.00	1.00	.83
	S2c2..	N/A	N/A	N/A	N/A	N/A	N/A	N/A
22								
	S1c1..	.00	.17	.17	.11	.33	.00	.33
	S1c2..	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S2....	.50	.17	.17	.28	1.00	1.00	.83
23								
	S2c1..	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S1....	.00	.00	.00	.00	1.00	1.00	.67
	S2c2..	N/A	N/A	N/A	N/A	N/A	N/A	N/A
24								
	S3....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S2....	.17	.33	.17	.22	.50	.83	.50
	S1....	N/A	N/A	N/A	N/A	N/A	N/A	.61
25								
	S1....	.33	.50	.50	.44	1.00	1.00	1.00
	S2....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
26								
	S1....	.00	.00	.17	.06	.83	.83	1.00
	S2....	.17	.17	.17	.17	1.00	1.00	1.00
27								
	S1....	.17	.50	.33	.33	.83	1.00	.94
	S2....	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 35 continued

Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim  
of Each Case

	GLOBAL I (INITIAL) RATINGS				GLOBAL 2 (FINAL) RATINGS			
	Lay	Police	Lawyer	Combined	Lay	Police	Lawyer	Combined
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN
<b>CASE</b>								
1								
	S1....	.17	.50	.83	.50	.83	.67	.83
	S2....	.33	.50	.50	.44	1.00	.67	.83
	O1....	.17	.17	.00	.11	.67	1.00	.50
								.72
2								
	S1....	.50	.83	.33	.56	1.00	1.00	1.00
	S2....	.33	.33	.00	.22	.67	.67	1.00
								.78
3								
	CS1...	.00	.00	.00	.00	.33	.00	.17
	CS2...	.00	.00	.00	.00	1.00	.83	.89
4								
	V1/S2.	.50	.67	.33	.50	1.00	1.00	.67
	S1....	.33	.67	.50	.50	1.00	1.00	1.00
5								
	S1....	.00	.00	.00	.00	.00	.00	.00
	S2....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6								
	S2....	.00	.00	.00	.00	1.00	1.00	.67
	S1....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
7								
	S1c1..	.00	.00	.17	.06	.83	1.00	.67
	S2c1..	.00	.00	.00	.00	.83	1.00	.83
	S1c2..	.67	.83	.67	.72	.50	.83	.50
	S2c2..	.00	.00	.00	.00	.83	1.00	.61
								.94

Table 35 continued

Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim  
of Each Case

	GLOBAL I (INITIAL) RATINGS				GLOBAL 2 (FINAL) RATINGS			
	Lay	Police	Lawyer	Combined	Lay	Police	Lawyer	Combined
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN
8								
	S2....	.67	.33	.67	.56	1.00	1.00	1.00
	S3....	.00	.17	.00	.06	.17	.50	.83
9								
	S1....	.17	.17	.67	.33	.50	.17	.33
	S2....	1.00	.50	.67	.72	1.00	.83	.89
10								
	S1....	1.00	.83	.67	.83	1.00	1.00	.67
	S5....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S4....	.00	.00	.00	.00	.67	.83	.50
	S3....	.17	.33	.17	.22	1.00	1.00	.67
11								
	S2....	.00	.00	.17	.06	1.00	1.00	.83
	S1....	.00	.00	.00	.00	1.00	1.00	1.00
12								
	V1....	.50	.50	.50	.50	1.00	.83	.83
	S2....	.17	.17	.00	.11	.17	.33	.00
	S1....	N/A	N/A	N/A	N/A	N/A	N/A	N/A
13								
	S2....	.17	.00	.33	.17	.17	.33	.17
	S5....	.00	.00	.00	.00	.67	.67	.72
	S1....	.00	.17	.33	.17	1.00	1.00	1.00
14								
	S1....	.17	.17	.33	.22	1.00	.83	1.00
	S2c1..	.17	.00	.00	.06	1.00	1.00	1.00
	S2c2..	.17	.33	.00	.17	1.00	1.00	.83

Table 35 continued

Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim  
of Each Case

	GLOBAL I (INITIAL) RATINGS				GLOBAL 2 (FINAL) RATINGS				
	Lay	Police	Lawyer	Combined	Lay	Police	Lawyer	Combined	
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	
15									
	S1....	.00	.00	.00	.00	.67	.67	.50	.61
	S2....	.33	.17	.33	.28	.83	.83	.67	.78
16									
	V2c1..	.00	.33	.17	.17	1.00	1.00	.83	.94
	S1c2..	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	V2c2..	.00	.17	.33	.17	.67	1.00	1.00	.89
17									
	V1c1..	.50	.50	1.00	.67	.83	1.00	.83	.89
	V1c2..	.50	.50	1.00	.67	.83	1.00	.83	.89
	S1....	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	S2....	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
18									
	S2....	.33	.00	.00	.11	.67	1.00	1.00	.89
	S1....	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
19									
	S3....	.33	.83	.50	.56	.33	.00	.33	.22
	S1....	.00	.00	.00	.00	1.00	1.00	1.00	1.00
	S2....	.00	.17	.00	.06	.33	.00	.00	.11
20									
	S1....	.17	.33	.50	.33	.33	.00	.17	.17
	S4....	.00	.00	.00	.00	.00	.00	.00	.00

Table 35 continued

Accuracy of Outcome: Global Ratings of Each Panel for Each Suspect/Victim  
of Each Case

	GLOBAL I (INITIAL) RATINGS				GLOBAL 2 (FINAL) RATINGS				
	Lay	Police	Lawyer	Combined	Lay	Police	Lawyer	Combined	
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN	
21									
	S1c1..	.50	.00	.50	.33	.33	.00	.17	.17
	S2c1..	.17	.17	.33	.22	.50	.67	1.00	.72
	S1c2..	.67	1.00	.67	.78	1.00	1.00	.83	.94
	S2c2..	.17	.17	.17	.17	.50	.67	1.00	.72
22									
	S1c1..	.00	.17	.17	.11	.33	.00	.33	.22
	S1c2..	.17	.17	.00	.11	1.00	.83	.67	.83
	S2....	.50	.17	.17	.28	1.00	1.00	.83	.94
23									
	S2c1..	.00	.00	.33	.11	1.00	1.00	1.00	1.00
	S1....	.00	.00	.00	.00	1.00	1.00	.67	.89
	S2c2..	.00	.00	.33	.11	.83	.33	.83	.67
24									
	S3....	.00	.17	.33	.17	.83	.83	1.00	.89
	S2....	.17	.33	.17	.22	.50	.83	.50	.61
	S1....	.00	.00	.00	.00	.67	.33	.17	.39
25									
	S1....	.33	.50	.50	.44	1.00	1.00	1.00	1.00
	S2....	.33	.50	.33	.39	.33	.33	.33	.33
26									
	S1....	.00	.00	.17	.06	.83	.83	1.00	.89
	S2....	.17	.17	.17	.17	1.00	1.00	1.00	1.00
27									
	S1....	.17	.50	.33	.33	.83	1.00	1.00	.94
	S2....	.17	.17	.17	.17	.33	.00	.83	.39

Outcome accuracy varied across suspects and across cases, ranging from 0 to 1.00. For the large majority of suspects, the "accuracy" of the Global 2 ratings were higher than the Global 1 ratings.

At the early stage of an investigation, little information is known or available for some suspects, giving rise to "undecided" ratings which are counted as "misses," thus leading to low hit rates for the initial ratings.

#### **What is the Overall Accuracy of the Ratings for Each Suspect for Each Panel?**

Table 35 also presents the accuracy of the overall Global ratings for each panel type (in terms of mean hit rate) for each suspect/victim in the case. For the most part, little variability in outcome accuracy of the ratings of suspects existed among the different types of panels. Greater differences in outcome accuracy of the ratings of suspects occurred between the Global 1 and Global 2 ratings with the second Global rating generally having much higher hit rates than the initial global rating. Thus, with a greater amount of case information, the hit rates are higher.

#### **What is the Overall Accuracy of the Ratings for Each Panel Member?**

Table 36 provides the accuracy of the overall Global Ratings for each panel member. The overall Global 1 (initial) hit rates varied from zero to 45%. The overall Global 2 (final) hit rates varied for the panel members from 50% to 80%. With the increased amount of information in the file to review for the final ratings, the panel members hit rates improved considerably between the initial and final ratings.

The use of a panel when the level of information in the file is minimal (as in the case of the Global ratings collected prior to the polygraph administration) is not recommended since the level of accuracy is so low.

Table 36

## Accuracy of Ratings for Each Panel Member

<u>Panel</u>	<u>Rater</u>	<u>Initial Hit Rate</u>			<u>Final Hit Rate</u>		
		<u>Guilty</u>	<u>Innocent</u>	<u>Overall</u>	<u>Guilty</u>	<u>Innocent</u>	<u>Overall</u>
Lay	1	31.4	0.0	27.5	85.7	40.0	80.0
	2	42.9	0.0	37.5	80.0	40.0	75.0
	3	22.9	0.0	20.0	77.1	40.0	72.5
	4	11.4	0.0	10.0	68.6	20.0	62.5
	5	20.0	0.0	17.5	82.9	40.0	77.5
	6	14.3	20.0	15.0	77.1	20.0	70.0
Police	1	22.9	0.0	20.0	80.0	60.0	77.5
	2	51.4	0.0	45.0	80.0	20.0	72.5
	3	17.1	20.0	17.5	77.1	60.0	75.0
	4	42.9	0.0	37.5	80.0	40.0	75.0
	5	22.9	20.0	22.5	85.7	0.0	75.0
	6	20.0	0.0	17.5	74.3	40.0	70.0
Lawyer	1	31.4	20.0	30.0	80.0	40.0	75.0
	2	31.4	0.0	27.5	82.9	60.0	80.0
	3	00.0	0.0	0.0	48.6	60.0	50.0
	4	20.0	0.0	17.5	74.3	40.0	70.0
	5	22.9	0.0	20.0	68.6	20.0	62.5
	6	25.7	0.0	22.5	71.4	40.0	67.5

Although the variability in overall final hit rates among panel members is relatively small, some data suggested additional data collection of personal characteristics of panel members might help identify which characteristics (from a selection viewpoint) might be valid in identifying "the highly accurate" panel members from "the less accurate" panelists. It is also possible that additional training of panel members might help in improving accuracy.

Table 36 also presents the hit rates for guilt and innocence for each panel member. For the final ratings, all panel members were more accurate in identifying guilty suspects than innocent ones.

### **What is the Overall Accuracy of the Specific Ratings?**

The accuracy of the specific (relevant question) ratings for each suspect/victim in each case is presented in Appendix V. Are ratings of specific questions more accurate than global ratings of guilt or innocence?

Table 37 presents the hit rates for each suspect in each case. The hit rate of the global rating is provided along with the highest hit rate among the set of specific questions rated. Thus, the accuracy of the global rating is compared to the rating of the best (i.e., most accurate) specific rating. This approach "stacks the deck" in favor of selecting the best specific items. The rationale for this approach would be to determine whether greater specificity in identifying ground truth would be more appropriate than a general assessment of ground truth. Perhaps it is easier to determine guilt or innocence (i.e., ground truth) relative to a specific question than to a global statement of guilt or innocence to committing a specific crime. The data in Table 37 indeed suggests that the best (in terms of highest accuracy rates) of specific relevant questions is more accurate than the global ratings of guilt or innocence. The use of specific questions and the measurement of accuracy in those terms seems a promising avenue to explore in future studies. It must be noted, however, that even though the accuracy corresponding to the highest specific ratings is higher than the global ratings, the hit rate associated with the initial rating are still very low, and would still suggest that panels not be used when the level of information in the file

Table 37

## Accuracy of Outcome: Global and Specific Ratings for Combined Sample

<u>Case</u>	<u>Suspect</u>	INITIAL RATINGS			FINAL RATINGS		
		<u>Global</u>	<u>Highest Specific</u>	<u>Specific Question #</u>	<u>Global</u>	<u>Highest Specific</u>	<u>Specific Question #</u>
1	S1	.06	.28	2	.11	.28	2
	S2	.44	.33	2	.83	.94	2
	O1	.06	.00	---	.06	.11	2
2	S1	.56	.33	2	1.00	1.00	1, 5, 6
	S2	N/A	N/A	---	N/A	N/A	---
3	CS1	N/A	N/A	---	N/A	N/A	---
	CS2	.00	.00	---	.89	.89	1
4	X1/S2	.50	.83	3	.89	1.00	1, 3, 4, 8, 9
	S1	.06	.56	10, 11	.00	.39	11, 12
5	S1	.00	.17	4, 5	.00	.11	5
	S2	.22	.89	3	.22	.89	3
6	S2	N/A	---	---	N/A	---	---
	S1	.00	.06	4	.89	1.00	4
7	S1C1	.06	.72	1	.83	.56	6, 7
	S2C1	N/A	---	---	N/A	---	---
	S1C2	.72	.72	1	.61	.56	6, 7
	S2C2	N/A	---	---	N/A	---	---

Table 37 continued

Accuracy of Outcome: Global and Specific Ratings for Combined Sample

<u>Case</u>	<u>Suspect</u>	INITIAL RATINGS			FINAL RATINGS		
		<u>Global</u>	<u>Highest Specific</u>	<u>Question Number(s)</u>	<u>Global</u>	<u>Highest Specific</u>	<u>Specific Question Numbers</u>
8	S2	.56	.67	6	1.00	1.00	2, 6, 7
	S3	N/A	---	---	N/A	---	---
9	S1	.56	.65	1	.67	.50	1
	S2	N/A	---	---	N/A	---	---
10	S1	.83	.50	5, 8	.89	.89	1, 5, 8
	S5	N/A	---	--	N/A	---	---
	S4	.00	.00	---	.67	.67	2, 8
	S3	.22	.50	33	.89	.76	8
11	S2	N/A	---	---	N/A	---	---
	S1	.00	.11	4	1.00	.94	3
12	V1	N/A	---	---	N/A	---	---
	S2	.11	.11	6	.17	.94	6
	S1	N/A	---	---	N/A	---	---
13	S2	.00	.83	1, 2, 3, 6, 7	.72	1.00	7.10
	S5	N/A	---	---	N/A	---	---
	S1	.17	.22	6	1.00	.94	1, 2, 4, 5, 6 7, 9
14	S1	N/A	---	---	N/A	---	---
	S2C1	.06	.11	1, 6	1.00	.89	4
	S2C2	.17	.11	1, 6	.94	.89	4

Table 37 continued

## Accuracy of Outcome: Global and Specific Ratings for Combined Sample

<u>Case</u>	<u>Suspect</u>	INITIAL RATINGS			FINAL RATINGS		
		<u>Global</u>	<u>Highest Specific</u>	<u>Specific Question #</u>	<u>Global</u>	<u>Highest Specific</u>	<u>Specific Question #</u>
15	S1	.00	.17	4	.61	.89	4
	S2	N/A	---	---	N/A	---	---
16	V2C1	N/A	---	---	N/A	---	---
	X1C2	.06	.11	6, 7	1.00	1.00	3, 4, 5, 6, 7
	V2C2	N/A	---	---	N/A	---	---
17	V1C1	N/A/	---	---	N/A	---	---
	V1C2	N/A	---	---	N/A	---	---
	S1	.67	.39	3, 4	1.00	.94	1, 3, 4
	S2	N/A	---	---	N/A	---	---
18	S2	N/A	---	---	N/A	---	---
	S1	.06	.17	3	.94	.89	1, 4, 8, 9
19	S3	.00	.72	2	.56	.67	4, 6
	S1	.00	.06	8			
	S2	.17	.00	---	.61	.00	---
20	S1	.33	.50	6	.17	.33	7.PA
	S4	N/	---	---	N/A	---	---

Table 37 continued

Accuracy of Outcome: Global and Specific Ratings for Combined Sample

<u>Case</u>	<u>Suspect</u>	INITIAL RATINGS			FINAL RATINGS		
		Highest Global	Specific Global	Question #	Global	Highest Specific	Specific Question #
21	S1C1	N/A	---	---	N/A	---	---
	S2C1	N/A	---	---	N/A	---	---
	S1C2	.78	.50	5	.94	.83	6
	S2C2	N/A	---	---	N/A	---	---
22	S1C1	.11	.44	4	.22	1.00	4
	S1C2	N/A	---	---	N/A	---	---
	S2	.28	.28	2	.94	1.00	2,3,5
23	S3	N/A	---	---	N/A	---	---
	S1	.00	.00	---	.89	1.00	5,7
	S2C2	N/A	---	---	N/A	---	---
24	S3	N/A	---	---	N/A	---	---
	S2	.22	.11	5,6,7,9 11,12	.61	.61	12
	S1	N/A	---	---	N/A	---	---
	S1	.44	.22	1	1.00	1.00	2
25	S1	N/A	---	---	N/A	---	---
	S2	N/A	---	---	N/A	---	---
26	S1	.06	.06	1,2,3,6	.89	.94	4
	S2	.17	.28	2	1.00	1.00	3,5,6
27	S1	.33	.33	4	.94	.94	4,6
	S2	N/A	---	---	N/A	---	---

is minimal.

The above findings also suggest that certain polygraph techniques that utilize specific relevant questions (e.g., Modified General Question Technique) might be validated using more specific measures of ground truth, than polygraph techniques (e.g., Zone of Comparison) that might best be validated with global measures of ground truth.

### **Comparison of Polygraph Results with Panel Ratings**

The ratings of the panel for each suspect/victim were compared to the polygraph results obtained from these same suspects/victims. When comparing the polygraph and panel rating results, a "hit" can occur in two ways:

1. The polygraph result is "Deception Indicated" (DI) and the panel rating is "Probably did commit the above crime" or "Definitely did commit the above crime," or
2. The polygraph result is "No Deception Indicated" (NDI) and the panel rating is "Probably did not commit the above crime" or "Definitely did not commit the above crime."

A number of cases had one or more suspects/victims with polygraph data but no confession data to verify the results. Some cases had statements of confession in the file which were unrelated to the crime committed. Polygraphed suspects/victims with no confession or with an unverifiable confessions (as indicated by a "0," "OA," or "OB" rating on the Quality of Confession scale when rated by the independent judges) were excluded from the hit rate analyses and were not counted as part of the total when the hit rates were calculated.

"Undecided" ratings made by panel members are counted as "misses."

Table 38 contains the hit rates of the polygraph results with the global panel ratings across all cases and all panel types. The overall hit rates for the initial (Global 1) rating was 63.15% and for the final (Global 2) ratings was 78.78%.

Table 38  
Hit Rates: Global Ratings of Panel with Polygraph Results by Panel Type

Global Rating	Panel Type				Combined
	Lay	Police	Lawyer		
Initial	56.52%	66.67%	66.67%		63.15%
Final	80.99%	78.32%	77.03%		78.78%

The hit rates for the respective panels varied only slightly. The initial (Global 1) hit rates were 56.52% (lay panel), 66.67% (police panel), and 66.67% (lawyer panel). The hit rates for the final (Global 2) panel ratings were higher than the initial ratings for all panels with hit rates of 80.99% (lay panel), 78.32% (police panel), and 77.03% (lawyer panel).

Table 39 provides the hit rates of the global ratings with the polygraph results for each panel member and for each case. The hit rates for the final (Global 2) ratings were substantially higher than the initial ratings for every panel member. The proportion of hits ranged from 0.063 to 0.429 for panel members for the initial ratings and ranged from 0.619 to 0.810 for the final ratings.

When each case is examined separately, it can be seen the hit rates vary across cases even though the final ratings are generally much higher than the initial ratings. The most dramatic change occurred in Case 11 where the initial hit rate was zero or near zero for all panels and the final hit rate was either 1.000 or 0.917 for the panels.

Table 40 presents the hit rates when comparing the polygraph results to the independent judges estimates of ground truth. The obtained hit rate was 75.7%. Additional details are provided in Appendix U.

A power analysis (refer to Table 41) was completed comparing the combined panels' final ratings of guilt and innocence with the independent judges ratings. For the final ratings, a rating of "definitely did commit" "probably did commit" was considered as "did commit crime." A rating of "definitely did not commit" or "probably did not commit" was considered as "did not commit". Final ratings of "undecided" were omitted from the power analysis. The odds ratio is 3.26 with a 95% confidence interval of 1.94 to 5.47 and Chi-Square value of 24.22 ( $p < .000001$ ).

A 3.26 odds ratio indicates that for those where the final ratings were "did commit," the proportion guilty to innocent (481/46) is 326% greater than the proportion guilty to innocent (109/34)

Table 39

Hit Rates: Global Ratings with Polygraph Results by Panel Member  
and by Case Number

	INITIAL RATINGS			FINAL RATINGS		
	Layperson	Police	Attorney	Layperson	Police	Attorney
	MEAN	MEAN	MEAN	MEAN	MEAN	MEAN
<b>Overall</b>						
Average	.206	.243	.262	.733	.717	.701
<b>Panel Member</b>						
1.....	.175	.317	.413	.698	.730	.746
2.....	.317	.429	.397	.746	.746	.778
3.....	.270	.143	.063	.698	.746	.619
4.....	.095	.333	.159	.651	.762	.714
5.....	.238	.159	.317	.810	.667	.635
6.....	.143	.079	.222	.794	.651	.714
<b>Case Number</b>						
1.....	.222	.389	.444	.833	.778	.722
2.....	.417	.583	.167	.833	.833	1.000
3.....	.000	.000	.000	.667	.417	.500
4.....	.417	.667	.417	1.000	1.000	.833
5.....	.000	.000	.000	.000	.000	.000
6.....	.000	.000	.000	1.000	1.000	.667
7.....	.167	.208	.208	.750	.958	.750
8.....	.333	.250	.333	.583	.750	.917
9.....	.583	.333	.667	.750	.500	.583
10.....	.389	.389	.278	.889	.944	.611
11.....	.000	.000	.083	1.000	1.000	.917
12.....	.333	.333	.250	.583	.583	.417
13.....	.056	.056	.222	.611	.667	.667
14.....	.167	.167	.111	1.000	.944	.944
15.....	.167	.083	.167	.750	.750	.583
16.....	.000	.250	.250	.833	1.000	.917
17.....	.500	.500	1.000	.833	1.000	.833
18.....	.333	.000	.000	.667	1.000	1.000
19.....	.111	.333	.167	.556	.333	.444
20.....	.083	.167	.250	.167	.000	.083
21.....	.375	.333	.417	.583	.583	.750
22.....	.222	.167	.111	.778	.611	.611
23.....	.000	.000	.222	.944	.778	.833
24.....	.056	.167	.167	.667	.667	.556
25.....	.333	.500	.417	.667	.667	.667
26.....	.083	.083	.167	.917	.917	1.000
27.....	.167	.333	.250	.583	.500	.917

Table 40

Hit Rates: Polygraph Results with Independent Judges Estimate of Ground Truth

Hits	Misses	Appendices
75.7%	24.3%	U-18, U-19

Table 41

## Power Analysis of Global 2 Ratings

## Independent Judges

<u>Panels Members Final Ratings</u>	<u>Guilty</u>	<u>Not Guilty</u>
Did Commit	481	46
Did Not Commit	109	34
	590	80
		670

## Analysis of Single Table

Odds Rating = 3.26 (1.94 &lt;OR&lt; 5.47)

cornfield 95% confidence limits for OR

Relative risk = 1.20 (1.09 &lt;RR&lt; 1.32)

Taylor Series 95% confidence limits for RR

Ignore relative risk if case control study

	<u>Chi-Squares</u>	<u>P-values</u>	
Uncorrected:	24.22	0.0000009	----
Mantel-Haenszel:	24.19	0.0000009	----
Yates corrected:	22.81	0.0000018	----

## \*\*\*Exact Confidence Limits\*\*\*

Exact Lower 95% Confidence Limit = 1.94

Odds Ratio = 3.26

Exact Upper 95% Confidence Limit = 5.47

where the final ratings were "did not commit." Using an arcsin transformation, the power of finding a statistically significant result is .93 at the .05 alpha level and .96 at the .01 Alpha level. However, another way to view this classification table is the proportion of times that a person who was guilty (ground truth) was judged to be so by the raters. This occurred 481 of 590 or 81.5%. Thus if a person is guilty most of the time he was rated as guilty (81%). Unfortunately, if a person were innocent (ground truth) he was found (rated) innocent only 42.5% of the time. This means that almost 58% of the time an innocent person was mistakenly found guilty.

## **EXECUTIVE SUMMARY AND RECOMMENDATIONS**

A major purpose of conducting a panel study is to establish an alternate measure of ground truth. The present study examined the feasibility of using the decisions of panels to establish ground truth. This was explored by varying the size and composition of the panel and the level of information provided it.

The present study has attempted to minimize some of the methodological problems found in previous studies, namely:

1. Panel studies to date are problematic in that it is unclear how much and what types and quality of data the panels had to work with; and
2. Previous panel studies have access to confessions elicited from only failed polygraph tests (i.e., DI-confirmed cases). These confessions are associated with an unrepresentative set of charts and are not independent of the outcome of the polygraph test.

Regarding the first problem, it is imperative that the case data presented to the panel be as complete and accurate as possible. Often the case data is abstracted prior to presentation to the panel. The process of abstracting files may contaminate the panel review process since the pertinent facts presented in the abstract have been "filtered" by some person's idiosyncratic views of what is and what is not important. In addition, of special concern is whether or not panels have access to confessions elicited following a failed polygraph test. To the extent they do have these confessions, the panel will most likely affirm the confession, and the distinctiveness of these two methods for establishing ground truth becomes blurred.

To address the above mentioned problems, the present study:

- a. controlled the amount of information provided to the panel. For example, the

"partial" level presents panel members with all case data placed in chronological order, up to the point in time the polygraph test was administered (excluding of course the polygraph charts and confessions). Thus, ratings obtained from panel members were not contaminated by knowledge of confession statements and

- b. provided all case information in its original form to the panel members. This avoided the "filtering" of information by an abstractor.

Regarding the second problem mentioned above, confession studies have been criticized because confessors are a select group. While this is true, a more serious problem with confession studies is that the polygraph charts associated with confessions represent a select group (Iacono, 1989; Iacono & Patrick, 1987, 1988). In the usual case, confessions are obtained when the examiner attempts to elicit them from a subject whose charts are indicative of deception. Thus, confessions will be obtained only from those who produced deceptive charts. Individuals who generate nondeceptive charts will not be asked to confess. While many of these people will be innocent, some will be guilty persons who erroneously produced truthful-looking charts. The charts associated with these "false negative" errors will never become part of a validity study when confessions are used to establish ground truth. Likewise, an unknown number of innocent people will erroneously produce deceptive charts. However, these individuals will generally not confess to an act they did not commit, and these "false positive" errors will never become a part of the validity study. Note that if some innocent people do confess following a failed polygraph test, this type of error is incorrectly counted as a hit in validity studies using confessions.

Thus, the major problem with confession studies is that almost all of the cases wherein the original examiner made an error are systematically excluded from the data sample. The only cases included are those where both the charts indicate deception and the examinee confesses. Since numerical scoring of polygraph charts is highly reliable, giving these charts of confession verified guilty and innocent subjects to other examiners to rescore blindly will of course confirm the original examiner's assessment. As a consequence, in confession studies, because cases where potential errors could have been made are eliminated, inaccurate estimates of polygraph validity

will be obtained. Accuracy rates for guilty individuals will be especially overly optimistic and should be close to 100% when numerical scoring is used even if the polygraph test had zero validity. Interpreting the data from innocent subjects will also be misleading for these same reasons.

It is seldom recognized in the polygraph literature that the arguments developed above also apply to panel studies in which the panels have access to confessions elicited following failed polygraph tests. A confession is likely to be persuasive evidence to a panel. However, because confessions will be associated with an unrepresentative set of charts, confessions elicited following failed examinations would not be presented to the panel if a legitimate estimate of polygraph validity is desired.

As with other reported studies, the preponderance of confession-verified cases in the present study collected from the military archives are DI-confirmed. Thus, for the most part, confessions are only obtained when deceptive charts are produced. In the present study, several design improvements were incorporated to minimize the problems of unrepresentativeness in the cases reviewed by the panels:

1. Only multiple-suspect case files were reviewed by our panels. The multiple-suspect files selected contained a greater mixture of DI- and NDI-confirmed cases. Thus, with multiple suspect cases, panel members are not limited to reviewing only DI-confirmed cases or confessions elicited from failed exams. For example, a rape case would have two individuals who have taken the polygraph test where only one of the them told the truth. Ratings were obtained from the panel members for every suspect/victim in the multiple-suspect case file.
2. The panel members in the present study were not aware of which of the multiple suspects rated also had confession data, which subjects rated had been administered the polygraph exam, nor were they aware of the number of suspects who were guilty or not guilty. With so many unknowns, panels members could not

conclude that most suspects were DI and make ratings accordingly.

Thus, the present panel studied avoided some of the methodological problems of previous research.

### **Data Collection Procedure**

The basic data collection procedure followed in the present study involved the following steps:

1. Obtain authorization to conduct the research.
2. Obtain authorization to obtain the computerized polygraph databases.
3. Select "promising" files based on the information contained in the computer databases.
4. Select an initial set of "promising" criminal investigative files based on the information contained in the databases. The initial set of "promising" files selected was to provide the researchers with a sampling of cases from different criminal categories and provide the investigators with a better idea of the types of procedures used, materials collected, and types of ground truth data that might be available (e.g., confessions, admissions, etc.) for different polygraph outcomes.
5. Each branch then pulled the requested files from its central repository, and "sanitized" the file following procedures we developed. The files were photocopied and the copies were "sanitized" by removing identifying information (i.e., names) and replacing it with our nomenclature: Suspect 1 (S1), Suspect 2 (S2), Victim 1 (V1), Witness 1 (W1), Other Person (O1), etc.
6. When we received the sanitized copies, we organized the file following the

procedure discussed earlier in this report. This process involved putting the contents of the file in chronological order, numbering the pages, placing markers to separate the file contents into two time periods (pre- and post-polygraph exam), and abstracting the case.

7. Based on the review of the initial set of cases, PDRI decided to select only multiple-suspect cases. Since the large majority of single-subject cases were DI-Confirmed, multiple-suspect cases were believed to be more "balanced" since some suspects would be innocent and others guilty and some subjects polygraphed would be DI and other suspects would be NDI. Most multiple-suspect cases were found in the rape, larceny, murder, and child abuse criminal categories.
8. Steps 4 and 5 were repeated again until an adequate number of files were retained.
9. The investigators reviewed the confession/admission statements and rated them on a "Quality of Confession" scale. Cases presented to the panel were confession-verified cases meeting our ground truth criteria.

#### **Review of Cases by Panel**

Twenty-seven confession-verified multiple suspect criminal investigative files were then selected and reviewed by three six-member panels consisting of lay persons, police investigators, and attorneys.

Panel members were presented with case material corresponding to two points in time. The level of information provided for each point in time varied. Time 1 materials contained the case investigative information collected *prior* to the first polygraph examination administered. Time 2 materials contained all investigative information (excluding polygraph results but including confessions) collected until the case was closed. Ratings were made on all of the material presented

in each time period.

Panel members rated the cases independently, making global and specific ratings regarding the probability of various subjects within each case having committed the crime or specific incident.

## Results

For the present study, measures of interrater reliability were calculated using Shroat and Fleiss' (1979) Model 2 for determining the Average Composite Reliability. The Average Composite Reliability is a measure of the reliability of a panel of raters and is the appropriate measure to report when a panel of more than one individual will be used to make the actual ratings in the future. The reliability of the specific ratings and of the global ratings for the two time periods were reported separately.

For all six-member panels, the mean reliability of the initial ratings of the specific questions for the 27 cases, averaged across the three panels is 0.182; and 0.625 for the final ratings of those same questions. The three types of panels had relatively similar reliability coefficients for the specific ratings over the two time periods. Since the reliability of the specific ratings was so low, particularly for the initial ratings, it is not recommended to use these ratings for future research. The initial ratings were gathered at Time 1 (just prior to the administration of the polygraph exam). The level of information in the case files at Time 1 is scant for some of the suspects, so it is difficult to make ratings on specific questions for suspects where little information exists. Rating suspects where little information exists, naturally will lead to unreliable results.

More encouraging results were found for the reliabilities of the global ratings. When all six members of each panel were included in the analysis the average composite Global 1 reliability is 0.71 for the lay panels, 0.72 for the police panel, and 0.81 for the attorney panel. The average Global 1 rating reliability across all groups is 0.75.

The reliability of the average Global 2 rating is 0.92 for the lay panel, 0.93 for the police panel,

and 0.92 for the attorney panel. The reliability of the average Global 2 rating across all groups is 0.93.

The reliabilities of the global ratings were also computed for each panel when the size of panel varied from two to six members. The larger the panel (i.e., the more people rating each case), the higher the overall or composite reliability.

Considering the importance of the decision the panel is being asked to make (i.e., guilt or innocence), the reliability coefficient should be as close to 1.00 as possible. We suggest a minimum reliability of at least 0.90. Regardless of panel size, none of the panels in the partial information (i.e. Global 1 Rating) condition achieved this minimum level of reliability. The highest level of reliability achieved for the Global 1 ratings was the six-member lawyer panel (reliability=0.806).

### **Accuracy of Outcome**

The panel ratings were compared to an alternate measure of ground truth (i.e., a verified confession). To determine whether or not a confession had been verified, the following procedure was employed:

1. Two judges independently completed the "Quality of Confession" Form to determine whether or not a suspect's confession was verified.
2. Both judges concurred independently that "the confession verified the accusation contained in the relevant question with almost certainty".

The confession-verification ground truth criterion used by the judges was more stringent than used by military personnel rating the original files (military personnel ratings of verification were obtained from the computer database). Some confessions were not rated as "verified" by the judges because of the following reasons:

- a. The confession was unrelated to the accusation, no way of confirming or disconfirming.
- b. The subject confessed to something other than what had been covered by the accusations in the polygraph questions.
- c. The subject confesses to something that may be covered by the accusations in the polygraph questions but can not be verified because of the terminology being used (e.g., is the "warehouse" mentioned in the confession the same as "Area 27" mentioned in the polygraph question?)

Accuracy of outcome was determined by comparing the panel ratings with the independent judges ground truth criteria. The overall accuracy of the Global 1 (initial) ratings was 22.5% for the combined panel of 18 members and 71.5% for the Global 2 (final) ratings. The accuracy of the overall global ratings across all cases for each panel was fairly similar across the different panels. The Global 2 "hit rates" were substantially larger than the Global 1 hit rates for all panels. The Global 2 hit rates as compared to the Global 1 hit rates were 72.9% to 21.3% (lay panel), 74.2% to 26.7% (police panel), and 67.5% to 19.6% (lawyer panel). Regardless of type of panel, the hit rates of the Global 1 ratings were small, indicating that panels, regardless of type, should not be used to establish ground truth at an early stage of the investigation (i.e., prior to the administration of the polygraph exam). The low hit rate for the Global 1 Rating is to a large extent due to the large percentage of "undecided" ratings made by panel members. When only partial information is provided to the panels, the level of information provided about the case is insufficient and panel members are to a large extent undecided about the guilt or innocence of the suspects. The less than perfect "hit rates" for the Global 2 ratings indicates that the confessions themselves were not seen as adequate evidence of ground truth in over 25% of the cases.

### **Summary of Panel Study Results**

The present research project examined the feasibility of using the decisions of panels to establish

ground truth. This was explored by varying the size and composition of the panel. Two types of ratings (global and specific), over two time periods, were collected from the panels. The time periods corresponded to the level of information provided in the case file. Time period 1 constituted a partial level of information condition and contained all information in the case file prior to the administration of the polygraph exam. The second time period constituted a complete level of information condition and contained all the information in the case file except the polygraph charts and the results of the polygraph. For each suspect in each of 27 cases, panel members made both global ratings (e.g., Did S1 commit the above crime?) and specific ratings corresponding to the relevant questions asked during the course of the polygraph exam (e.g., Did you ever violently strike V1 in the head area?).

The primary dependent variable of the study was an alternate measure of ground truth, the assessment of guilt/innocence by a verified confession. The two principal investigators of the present study, using a "Quality of Confession" scale, independently rated the confessions in each case. Only those cases where both investigators agreed that "the confession verified the accusation contained in the relevant question with almost certainty" were given to the panels. The investigators' determination of a confession's verification was more stringent than the military's branch's designation of verification (as indicated in the computer files). A number of cases previously confirmed by the military were not "confirmed" using the criteria used by the investigations in verifying confessions. The panel ratings were compared to the alternate ground truth measure and the results, measuring the accuracy of outcome, were reported in terms of hit rates. The alternate ground truth measure was the independent judge's verification of the confession. The average composite reliability of the panel ratings and a comparison of the panel ratings with the results of the polygraph were also reported. Table 42 summarizes the major findings of the study.

#### **Did the Type of Panel Make a Difference?**

Very few differences were found among the three types of panels (lay, police, and attorney) in terms of reliability, accuracy of outcome, and comparisons with polygraph results. A primary

difference was the cost required to hire the panel members to review the cases.

In terms of cost considerations for the present study, the attorney panel members were paid 3.87 times more than the lay panel members and 2.69 times more than the police panel members. The police panel members were paid 1.44 times more than the lay panel. The differences in reliability between lay, police, and lawyers on the final ratings are so small given the huge differences in cost for these 3 types of raters, a cost/benefit analysis would recommend using lay people as raters if reliability is the only consideration.

### **What is the Optimal Size for a Panel?**

Assuming we wish to maintain reliability of at least 0.90, then we would not recommend using panels of any size for the partial information (i.e., Global 1 Rating) condition.

Table 42

## Overall Findings

Statistic	Time Period	Type of Rating	Table	Panel Composition			Combined Average
				Lay	Police	Lawyer	
Average Composite Reliability	1	Global 1	24	.711	.723	.806	.747
	2	Global 2	24	.919	.934	.924	.926
	1	Specific (Initial)	22	.178	.106	.261	.182
	2	Specific (Final)	22	.626	.617	.632	.625
Accuracy of Outcome (as compared to independent judges' assessment of confession verification)	1	Global 1	34	21.3%	26.7%	19.6%	22.5%
	2	Global 2	34	72.9%	74.2%	67.5%	71.5%
	1	Specific (Initial)	37	Accuracy of outcome is provided for each suspect for each case in Appendix V			
	2	Specific (Final)	37				
Polygraph Result Hit Rate	1	Global 1	38	56.5%	66.7%	66.7%	63.2%
	2	Global 2	38	81.0%	78.3%	77.0%	78.8%
	1	Specific (Initial)	39	20.6%	24.3%	26.2%	23.7%
	2	Specific (Final)	39	73.3%	71.7%	70.1%	71.7%

### **Are Confessions Suitable for Establishing Ground Truth?**

An unexpected finding was that even though the cases presented to panelists were highly selected and prescreened such that they were believed to contain confessions that involved admissions to questions covered in the polygraph tests, the panelists rejected suspect confessions in over 25% of the cases. This result indicates that confessions themselves are likely to have questionable validity as indices of ground truth.

### **Will Better Selection or Training of Panel Members Produce "Better" Results?**

The overall Global 1 (initial) hit rates of the individual panel members varied from zero to 45%. The overall Global 2 (final) hit rates varied for the panel members from 50% to 80%. Although the variability in overall hit rates among panel members was relatively small, it is possible that the collection of data regarding personal characteristics of panel members might help identify characteristics that, from a selection viewpoint, might be useful for separating "the highly accurate" panel members from "the less accurate" panelists. It is also possible that additional training might help improve accuracy.

### **Recommendations for Future Research**

With the full understanding that the requirements of our project placed additional demands on each branch's workload, we are very appreciative of the high level of cooperation given by each branch in helping us carry out the above data collection procedure. We are also very appreciative

of the Polygraph Institute's assistance in providing the additional human resources available to help sanitize the files.

#### **A. Use of Military Records**

##### **1. Access to Data Files**

In future studies, where collection of archival data is required, we strongly recommend that researchers be given full access to the actual criminal investigative files. If this access requires additional security clearances, then we recommend that such clearances be arranged.

Clearly the most time consuming and costly step of the data collection procedure is the "sanitization" of files. If the investigators had access to the files in the central repository of each military branch, a more judicious selection of case files could have been made. Instead of relying solely on the information contained in the computer databases for case selection, the investigators could have used the computer database for only the initial screening of cases. Then, prior to any sanitization of files, the investigators could have reviewed the actual case files and determined whether or not the criterion for ground truth (i.e., "Quality of Confession") was met. Unfortunately, in the present study a number of files were sanitized and abstracted, which did not meet our criterion for ground truth and which could not be used for the panel study. If full access to the files were permitted, we could have reduced the number of files requested, and likewise reduced the time and costs associated with sanitizing and abstracting those files.

## **2. Standardization of Military Polygraph Procedure**

The detailed procedures for the use of the polygraph in the military have been presented in the text and appendices of this report. For each branch, information is provided regarding the polygraph's applicability, uses, authorization for approval, authorization forms and criteria, administration procedure, equipment maintenance, reporting method, routing records, and supervision and review. For the most part, the procedures are similar across military branches even though the level of specificity in discussing the information mentioned above varies widely. The Air Force provides the greatest amount of detail.

Although it probably would be difficult to implement, it would be highly desirable to have a common set of procedures to be followed by all branches. From the perspective of a future validation research design, it would be important to clarify and come to some consensus on the "Supervision and Review" section of each branch's documented procedures. In particular, the quality review and control functions need to be more explicitly stated, particularly how discrepancies are resolved between the original examiner and the reviewer.

Currently, the Navy/Marines procedures dictate that a supervisory official who is a certified polygrapher will exercise professional and technical supervision. There is no mention about quality control procedures. The Army procedures discuss how each set of charts is reviewed by supervisors. The reviewers may directly contact the original examiner, and if the two do not interpret

the charts in the same way, the supervisor may require the exam to be repeated. For polygraphs found to be "inconclusive" or "no opinion," a repeat exam may be administered without having to request approval again (it's unclear if this applies to the original judgment, to the reviewer's judgment, or to both). The Air Force procedures specify that each case is reviewed by a quality control supervisor who prepares a written critique of each examination and passes that information along to the original examiner and headquarters (HQ). Based on this critique, HQ decides whether the exam should be repeated or not. The Air Force procedures also specify that for counterintelligence security polygraph exams, a second examiner must be *present* during the examination and must conduct an immediate on-scene numerical evaluation and quality control review after the first examiner conducts the exam and scores the chart.

Each branch needs to address in its procedures the following quality control issues:

1. The type of contact, if any, between the original examiner and the reviewer.
2. The manner by which "scoring" errors are corrected, and how the original polygraph score or result is amended.
3. In cases other than those with "inconclusive" or "no opinion" results under what circumstances is a repeat exam required and/or the original polygraph exam invalidated?

We recommend:

- a) no contact between original polygraph examiner and reviewer during the review process,
- b) all discrepancies need to be documented in writing with reviewer comments and placed in the case files,
- c) when the review process is completed, the reviewer may provide verbal feedback to the original examiner to avoid errors in the future,
- d) in cases other than those with "no opinion" or "inconclusive" results, repeat exams should only be required if the scoring and interpretation of the original polygraph exam has been invalidated (i.e., documented to be incorrectly administered or scored).

### **3. Standardization of Computer Databases**

For the present study, the initial selection of cases was determined from information contained in the polygraph computer databases from each branch. However, each branch stores different information from the case files and polygraph charts in its computer database. Although the present study was able to select cases from a core set of variables common to all databases (e.g., polygraph result, confirmation status), future research may wish to select cases using other variables which may be available in computer files for some branches but not others. Even when the data are available, it may be categorized differently. For example, "type of crime" was a relevant selection variable in the study but each branch had a different method for organizing the

criminal categories and assigned different numerical values to these categories. The present investigators had to create a new taxonomy of crimes with a common classification scheme so that similar crimes could be selected across branches.

Additionally, each branch uses a different software database to store and retrieve the polygraph information. The lack of a centralized database across all branches significantly hampered the present study's data collection efforts.

Since different types of data are collected and then stored in different databases using different software, this lack of standardization will pose significant obstacles to future researchers working with polygraph archival data.

It is a strong recommendation of this report to establish a centralized database using standard formats for data retrieval and storage. Likewise, standard types of information should be collected from all branches (refer to Table 2 in text).

The Polygraph Institute would be a logical choice to serve as the central repository for the computer files. Each branch would still maintain its own data, but a duplicate set of data with regular updates should be sent to the central repository. The standard types of information collected, and standard formats for data retrieval and storage could be finalized through a joint committee rep-

resenting all branches. In the meantime, the present data sets (in ASCII format) could be sent to a centralized repository of polygraph data.

## **B. Future Field Validity Studies**

To recapitulate the major findings of the panel study, the global and specific ratings of suspects' guilt or innocence that had been collected from the panels at Time 1 were found to be unreliable and insufficiently valid. Thus, panel ratings should not be used to establish ground truth. We must conclude the panel approach is unlikely to provide a valid measure of ground truth even when files are highly selected.

We believe there are two possible approaches for continued research in developing a proposed field validation research design: 1) redo the panel study, pushing the limits of available technology and following a number of additional specific recommendations for improving the design; or 2) consider a different approach to thinking about validity.

### **Redo the Panel Study**

We believe there is limited utility in extending the present study methodology to see if panels can be used to establish ground truth. If, however, this study is extended, we would recommend several changes.

**1) Organize Case Files Differently.** At present, when a summary report or Report of Investigation is included in the criminal investigative file, the report sometimes contains information collected before the polygraph examination that, however, does not appear in the file until after the polygraph test is administered. In the present study, this information, even though it was collected independent of the polygraph test outcome, was not presented to the panelists at Time 1. For example, in Case 1 of the panel study, the results of an important lab test were received after the polygraph exam had been completed. The lab results in this case provided critical factual information which ruled out the suspect's claim of accidentally dropping the baby due to falling asleep on the rocking chair. The lab results clearly ruled out this accidental cause and strongly suggested child abuse had occurred due to choking. Without physically cutting up the report, it is difficult to separate this information into the pre-polygraph (partial level of information) condition and post-polygraph (complete level of information) condition. The present study did not "cut-up" the summary report into its pre- and post-polygraph results. We did not take this labor intensive step and believe that it is unlikely, had we done so, that this study would have produced different results. However, this is really an empirical question, one that we did not examine directly.

**2) Obtain Better Confession Data.** The polygraph file should always have a signed admission or confession statement rather than a summary statement that a confession had been made, as was found in some cases. To help ensure the confession is verifiable as a measure of ground

truth, the confession should include an unambiguous admission of guilt to a specific crime and provide as many specific details as possible. Including more specific details in the confession/admission statement will help provide additional evidence and greater certainty in verifying the accuracy of the relevant questions asked during the polygraph exam. We also believe when a confession provides more specific data, it becomes more likely to hold up if the case goes to court.

We have been told by several military sources that if we require too many details in a confession statement, we may end up with unsigned confessions. Although we fully agree with the utility of having signed confessions and would not want to jeopardize this, it does raise the question as to why suspects who have already admitted to committing the crime won't admit to the specifics of the crime. We simply want to help ensure that if the specifics of the crime were not included in the confession, it was due to a suspect's reluctance to sign rather than a simple omission by the examiner to ask.

Occasionally, in the course of an investigation the crime has changed. For example, a case may open as a "rape" case and close with the crime classified as a "false statement". In cases such as these, the polygraph is often administered under the assumption that the suspect was a victim of "Crime 1" (e.g., the rape) but by the end of the investigation, the suspect is accused of and confesses to "Crime 2" (e.g., the false statement). Although the polygraph exam and all relevant

questions are directed to "Crime 1", the case is closed as being "Confirmed" when the admissions deal with a related but technically different issue than that addressed by the polygraph tests. These types of complex cases are probably not ideally suited for inclusion in polygraph validity studies that use panels to establish ground truth.

**3. Don't Allow Panelists to "Sit on the Fence."** The global rating scales used in the present study allowed panel members to select a middle "undecided" rating when guilt or innocence could not be determined. Rather than commit to a "guilty" or "innocent" verdict, often panel members, particularly at Time 1, assigned an "undecided" rating. The high number of "undecided" ratings led to low hit rates when panel ratings were compared to our measure of ground truth. It is uncertain to what extent our results were affected by the options given to the panel on our rating scale. We may have made it too easy for panelists to avoid making decisions that, given the evidence, may nevertheless have been appropriate. If this study were redone, we would recommend eliminating this decision alternative.

**4. Select Judges Because they are Prescreened to be Good Jurists..** Some panelists are likely to be better than others. The accuracy of the overall ratings varied considerably across raters. For example, Lawyer Number 3 had zero percent accuracy for the initial ratings and only 50% accuracy for the final set of ratings. Recall the final set of ratings were made after all case information including confessions were revealed. In contrast, Police Officer number 2 had a

45% accuracy for the initial ratings. Likewise, Layperson Number 1 and Lawyer Number 2 had 80% accuracy for the final ratings. Future studies should prescreen potential panel members to select the more accurate ones. Panelists could actually be pretested by having them review say 10 cases. We would then select only those panelists who "accurately decided" a high proportion of those 10 cases.

**5. Train Panelists How to Make Decisions.** Another method for improving the accuracy of the panels is to provide training. For example, one could review cases with panelists to teach them how to decide evidence in pilot or test cases that they review before reviewing the "real" cases for the study.

**6. Convene the Panel as a Group and Discuss Cases.** The present study had panel members working and rating cases independently, without any input or discussion from other panel members. Had we used the panel like an actual jury, fostering discussion and clarifying the facts and evidence of the case, a better panel outcome may have resulted.

**7. Use Panel Members Other than Defense Attorneys.** In addition to cost considerations, other evidence exists that suggest defense attorneys may not be the best panel members. Based on analysis of variance results, the attorneys were significantly more likely than the other two panels to rate criterion "guilty" suspects as not committing the crime . Issues of selection and

previous training of the lawyers as mentioned above may also play a role. Perhaps attorneys use different criteria for evaluating confession data than the other panels. Perhaps at issue is not whether the confession indicates "ground truth" guilt but rather "If my client made this confession, could I still avoid a guilty verdict in a court of law". This is somewhat speculative and requires further investigation but perhaps defense attorneys evaluate confession data in terms of how well it holds up in a court of law.

### **New Designs for Field Validation of the Polygraph**

Two new designs are proposed for the field validation of the polygraph. The first design (Design 1) discussed will involve collecting new criminal investigative cases across military branches. The second design (Design 2) can be implemented immediately and uses archival data. If Design 2 is not implemented immediately, then the previous recommendations regarding the standardization and centralization of databases should be implemented prior to starting Design 2.

As has been noted previously, a major problem with field validity studies centers on the use of ground truth data (confessions) that arise as a consequence of failing a polygraph test. To get around this problem, a field study should be conducted in which polygraph tests are administered but (1) the results of the test are not released, and (2) failed tests are not followed-up by post-test attempts to elicit confessions. There are many ways this objective could be accomplished; we recommend one below.

## **Design 1**

### **Basic Steps:**

1. For each military branch, select test sites in locations where a minimum of two polygraph examiners are always present.
2. Select, say, every fifth new criminal case opened at this site to be included in the field validation study. Although multiple suspect cases are more desirable, partly because of the greater likelihood of finding some innocent suspects, single-suspect cases can still be used. The time required from case opening to closing is generally much longer for multiple suspect cases and therefore a mix of single-suspect cases would be desirable. However, urinalysis cases or cases where subjects request polygraph exams for exculpatory purposes should not be included.
3. Obtain permission from the suspects to administer the exam, but explain that the exam is only being used as an investigative aid and, as a consequence, its results will not be revealed. Administer the polygraph exam to all preliminary suspects in the case. All case information will be provided to the examiner. The focus of the Design 1 will be to employ only the "Zone of Comparison (ZOC)" and "Modified General Question Technique (MGQT)" polygraph techniques. These techniques are used extensively. For example, the Army which has the largest polygraph database, the ZOC is used in approximately 42.4% of the criminal cases. The MGQT is used in approximately 39.7% of the cases. These techniques were chosen not only because of the frequency of their use but also because of the varying scoring schemes used. The ZOC makes use of a more "global" scoring

method whereas the MGQT employees a more "specific" method of scoring relevant questions.

4. Do not release the polygraph results to the suspects. Do not extract a confession or admission at this time. Do not discuss the results with anyone, including other polygraph examiners. Simply score the charts and place them in a sealed envelope. The sealed envelope should be sent to Headquarters and remain unopened until the case is closed. The case file and corresponding computer database should simply note for the record the existence of the "research"-only polygraph report.
5. If, in the course of the investigation, it becomes appropriate to administer a polygraph in the usual way, a second polygraph exam will be administered by a different investigator. The need for a second polygraph test would probably arise in only a very small proportion of cases. The second examiner will carry out the usual polygraph procedures. The suspect could be told that the second test has a different purpose from the first test and though similar will be a different test with different questions. Thus, the second examiner may collect pre- or post-test admission statements or confessions and proceed with the case as he/she normally would.
- 6) When the case is closed, if a confession or admission statement was obtained, the envelope can be opened by an independent group of researchers. The polygraph charts should be duplicated and sent to Quality Control for blind scoring to ensure the numerical scoring is correct and the results of the polygraph are valid.
- 7) While the charts are in Quality Control for review, a panel of five lay members will be provided the actual confession/admission statements. The panel members will independently rate the confession on the "Quality of Confession" Form used

in the present study. If a panel consensus regarding the validity of the confession is achieved, the confession will be viewed as verified and used as the measure of ground truth. This important step is necessary because, as the panel study showed, confessions/ admissions do not necessarily provide unequivocal documentation of ground truth.

- 8) When the results of the Polygraph are returned and verified by Quality Control, the polygraph results are compared to the panel's decision of ground truth.
- 9) The data collection efforts can stop when 100 cases with ground truth data for guilt and innocence are collected from each polygraph technique.

The major advantage of Design 1 is that it controls for the fact that polygraph results (from the first exam) are collected independently, without influencing the course of the investigation, and early in the course of the investigation prior to the extensive development of the case.

## **Design 2**

This design represents an extension of the Patrick and Iacono (1991) field study. A major problem with the Patrick and Iacono study was that there were an insufficient number of confirmed guilty people where the confirmation was independent of polygraph test outcome.

The second design makes use of archival data and can be implemented immediately.

Basic Steps:

- 1) A policy needs to be in place which will allow "unresolved" cases to remain open for a period of time, say six months, over and above the normal closing of the case that would result once all polygraph tests have been administered and no confessions are forthcoming. Examining the archival data supplied to us by the military, there are many such unresolved cases.
- 2) During this additional time period, a different investigator, with a fresh perspective and new enthusiasm, is assigned to the case to continue the investigation. The second investigator would not be informed of the previous polygraph results (i.e., he/she would be told the polygraph didn't resolve the case or turn up new leads) and will continue to follow up whatever new leads develop. Individuals who have been tested, regardless of test outcome (which, of course, they would be informed of) should be told there is no resolution to the case and the case will remain open. In a small but very important fraction of these "reopened" cases additional ground truth information will be collected independent of polygraph data.
- 3) The hope of this new approach is to locate the difficult-to-find cases where a measure of ground truth is established independent of polygraph outcome. For example, a case of petty theft where a maid had been accused of stealing private property from an officer may have closed unresolved. If the case were continued to be opened for a few more months, it may be that the apparent "stolen" property resurfaced and that either someone else stole the property and confessed when arrested for another offense or the merchandise was simply misplaced and no crime was committed in the first place. In this example, the maid would have been innocent of the crime, her name cleared, with an alternate measure of ground truth support (i.e., another's confession or the officer's admission that the property was misplaced).

A panel is needed here, too, to decide the adequacy of the ground truth information.

Design 2 has the advantage of maintaining the independence of polygraph results from the collection of additional ground truth data since ideally the new investigator will not have knowledge of the polygraph results when he/she continues the investigation. Because the new investigator could learn of the polygraph outcome anyway (e.g., because the results are divulged by those who took polygraphs and who spoke with the investigator), it would be useful to interview (by phone, letter, in person) the investigator to determine whether he learned of the polygraph outcomes. Even if he did, however, it would still leave the eventually collected ground truth data largely independent of the polygraph test outcome.

## References

- Abrams, S. (1977). *A polygraph handbook for attorneys*. Lexington, MA: Lexington Books.
- Abrams, S. (1989). *The complete polygraph handbook*. Lexington, MA: Lexington books.
- Barland, G. H. (1988). *Polygraph research guidelines: Initial draft*. Department of Defense Polygraph Institute. Unpublished manuscript.
- Barland, G. H. & Raskin, D. C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology, 12*, 321-330.
- Barland, G. H. & Raskin, D. C. (1976). *Validity and reliability of polygraph examinations of criminal suspects*. (Report No. 76-1, Contract No. N1-99-0001). Washington, DC: National Institute of Justice, Department of Justice.
- Bersh, P. J. (1969). A validation study of polygraph examiner judgments. *Journal of Applied Psychology, 53*, 399-403.
- Bradley, M. T. & Ainsworth, D. (1984). Alcohol and the psychophysiological detection of deception. *Psychophysiology, 21*, 63-71.
- Davidson, P. O. (1968). Validity of the guilty-knowledge technique: The effects of motivation. *Journal of Applied Psychology, 52*, 62-65.
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *Journal of Applied Psychology, 75*, 521-529.
- Elaad, E. & Kleiner, M. (1986). The stimulation test in polygraph field examinations: A case study. *Journal of Police Science and Administration, 14(4)*, 328-333.

- Fleiss, J. L., & Shrout, P. E. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 422-428.
- Ginton, A., Daie, N., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real life situation. *Journal of Applied Psychology, 67*, 131-137.
- Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Research in Personality, 19*, 373-385.
- Horvath, F. S. (1977). The effects of selected variables on interpretation of polygraph records. *Journal of Applied Psychology, 62*, 127-136.
- Horvath, F. S. & Reid, J. E. (1971). The reliability of polygraph examiner diagnosis of truth and deception. *Journal of Criminal Law, Criminology, and Police Science, 62*, 276-281.
- Hunter, F. L. & Ash, P. (1973). The accuracy and consistency of polygraph examiners' diagnoses. *Journal of Police Science and Administration, 1*, 370-375.
- Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests? In P.K. Ackles, J.R. Jennings, and M.G.H. Coles (Eds.) *Advances in Psychophysiology* (pp. 201-207).
- Iacono, W. G., Cerri, A. M., Patrick, C. J., & Fleming, J. A. (1992). Use of antianxiety drugs as countermeasures in the detection of guilty knowledge. *Journal of Applied Psychology, 77*, 60-64.
- Iacono, W. G. & Patrick, C. J. (1987). What psychologists should know about lie detection. *Handbook of forensic psychology*, 460-489.

- Iacono, W. G. & Patrick, C. J. (1988). Polygraph techniques. In R. Rogers Ed.), *Clinical assessment of malingering and deception*. New York: Guilford press.
- Iacono, W. G. (1989). *Can we determine the accuracy of polygraph tests?* In P. K. Ackles, J. R. Jennings, and M. G. H. Coles (eds.) Advances in Psychophysiology, Volume 4. Greenwich, CT: JAI Press.
- Janisse, M. P. & Bradley, M. T. (1980). Deception, information, and the pupillary response. *Perceptual and Motor Skills, 50*, 748-750.
- Lykken, D. T. (1979). The detection of deception. *Psychological Bulletin, 56*, 47-53.
- Lykken, D. T. (1981). *A tremor in the blood: Uses and abuses of the lie detector*. New York: McGraw Hill.
- Lykken, D. T. (1988). Detection of guilty knowledge: A comment on Forman and McCauley. *Journal of Applied Psychology, 73*, 303-304.
- Lykken, D. T. (July, 1990). Why (some) Americans believe in the lie detector while others believe in the guilty knowledge test. *Paper delivered at the International Congress of Psychophysiology, Budapest, Hungary*.
- Kircher, J. C., Horowitz, S. W., & Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior, 12*, 79-90.
- Office of Technology Assessment. (1983). Scientific validity of polygraphic testing: A research review and evaluation.

- Patrick, C. J. & Iacono, W. G. (1989). Psychopathy, threat, and polygraph test accuracy. *Journal of Applied Psychology, 74*, 347-355.
- Patrick, C. J. & Iacono, W. G. (1991). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology, 76*, 229-238.
- Podlesney, J. A. & Raskin, D. C. (1977). Physiological measures and the detection of deception. *Psychological Bulletin, 84*, 782-799.
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.
- Slowick, S. & Buckley, J. (1975). Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure, and GSR recordings. *Journal of Police Science and Administration, 3*, 305-309.
- Waid, W. M. & Orne, E. C. (1980). Individual differences in electrodermal lability and the detection of information and deception. *Journal of Applied Psychology, 65*, 1-8.
- Waid, W. M., Orne, E. C., & Wilson, S. K. (1979). Effects of level of socialization on electrodermal detection of deception. *Psychophysiology, 16*, 15-22.
- Wicklander, D. & Hunter, F. (1975). The influence of auxiliary sources of information in polygraph diagnosis. *Journal of Police Science and Administration, 3*, 405-409.